MANUAL OF MENTAL AND PHYSICAL TESTS

A BOOK OF DIRECTIONS COMPILED
WITH SPECIAL REFERENCE TO THE EXPERIMENTAL
STUDY OF SCHOOL CHILDREN IN THE
LABORATORY OR CLASSROOM

BY

GUY MONTROSE WHIPPLE, PH.D.
Assistant Professor of the Science and Art of Education, Cornell University
Author of "A Guide to High-School Observation," "Questions in General and Educational Psychology," "Questions in School Hygiene."

WARWICK & YORK,
INC.
BALTIMORE, U.S.A.
1910

MANUAL OF MENTAL AND PHYSICAL TESTS

PREFACE

Hitherto the literature of mental and physical tests has been scattered in numerous journals; the results obtained by different investigators have too often not been compared; indeed, in many cases where the methods have been divergent, comparison has been impossible. In consequence, there have been no recognized standards of procedure and none of performance. Nevertheless, I believe that the time has now come for the taking of an account of stock, and for the systematization of the available materials. This conviction, which is the outgrowth of my own interest in the experimental study of mental capacities, an interest that has been with me during the past ten years, has been confirmed by many suggestions from colleagues and friends, who have pointed out that a manual of directions for mental tests would meet a real need, and might further the cause of investigation. More particularly, at the instigation of Mr. C. H. Stoelting, of Chicago, who has undertaken to supply the apparatus and materials prescribed in this volume, I began, in March, 1906, to prepare a small handbook of mental tests. The impossibility of adequate treatment of the subject in small compass has, however, necessitated the expansion of that early undertaking into the present work.

In the introductory sections of the volume, I have sought t
show the general purposes of mental tests, to lay down rules
their conduct, and to explain the methods of treating data.
this connection I discuss the calculation of measures of g
tendency, measures of variability, indexes of correlatio
other statistical constants.

In the body of the volume, I have brought together, f
treatment, some fifty of the most promising tests. In
my plan has been to sketch the development of the
scribe a standard form of apparatus and method of

explain the treatment of the data secured, and to set forth the results and conclusions thus far obtained.

The tests that I have selected may not prove, ultimately, to be those of most value, but they are, I think, numerous enough, and varied enough in type, to furnish a working basis for investigations for some time to come.

In the choice of materials and methods, I have sought to follow a middle course; on the one hand to avoid the use of costly instruments of precision and of the elaborate methodology of the psychological laboratory, and on the other hand, to avoid the inexactness of make-shift apparatus and the unreliability of casual, unsystematic observation. My idea has been to supplement the exposition of the standard apparatus and method of procedure by suggestions for variations of apparatus or of method, so that each test will be carefully standardized, yet will retain a sufficient degree of flexibility.

Doubtless, to some readers, the instructions for the conduct of the tests will seem unnecessarily lengthy and detailed; but experience has convinced me that faulty results are to be traced, in quite the majority of instances, to the neglect of some seemingly trivial detail in the arrangement of the experimental conditions; so that instructions can scarcely be made too explicit in a manual of directions in which standardization is the object.

In explaining the treatment of data, my aim has been to make clear the arithmetic of the various formulas, without insisting, in every case, upon acquaintance with the mathematical reasoning upon which the formula is based.

And when I speak of "the results and conclusions thus far obtained," I speak with the intent to make clear what, I am sure, made evident more than once in the text, that this book presents, a closed chapter in the experimental investigation of mental ity, but a program of work to be done.

cknowledgments for aid should be numerous and ungrudg- ese have been made in part in the text, but in many in- aterial assistance has, perforce, gone without explicit ment. I wish, however, to make clear my indebted- Stoelting Co., for the loan of numerous cuts, to Dr.

Guy L. Noyes, of the University of Missouri, for assistance in the tests of vision, to Dr. H. H. Goddard, of Vineland, N. J., for the adaptation of the Binet-Simon tests to American conditions, to my colleague, Professor I. M. Bentley, as well as to my wife and to my mother, for the reading of proof, and to my colleagues, Professors Charles DeGarmo and E. B. Titchener, for almost daily advice and encouragement.

The inscription of the book to Professor Titchener is in token of my special debt to him as a teacher and as an expositor of the scientific method of attack in the solution of the problems of mental life.

GUY MONTROSE WHIPPLE.

Cornell University,
June, 1910.

TABLE OF CONTENTS

INTRODUCTORY

CHAPTER I.

CHAPTER II.

CHAPTER III.

THE TESTS

CHAPTER IV. ANTHROPOMETRIC TESTS

CHAPTER V. TESTS OF PHYSICAL AND MOTOR CAPACITY

CHAPTER VI. TESTS OF SENSORY CAPACITY.

CHAPTER VII. TESTS OF ATTENTION AND PERCEPTION

CHAPTER VIII. TESTS OF DESCRIPTION AND REPORT

CHAPTER IX. TESTS OF ASSOCIATION, LEARNING, AND MEMORY

CHAPTER X. TESTS OF SUGGESTIBILITY

CHAPTER XI. TESTS OF IMAGINATION AND INVENTION

CHAPTER XII. TESTS OF INTELLECTUAL EQUIPMENT

CHAPTER XIII. SERIAL GRADED TESTS FOR DEVELOPMENTAL DIAGNOSIS

INDEX OF TABLES

LIST OF ABBREVIATIONS

A. G. P.	Archiv für die gesammte Psychologie.
A. J. P.	American Journal of Psychology.
A. P.	Année psychologique.
Ar. P.	Archives de Psychologie.
B. J. P.	British Journal of Psychology.
C. C.	Columbia University Contributions to Philosophy, Psychology, and Education.
E. P.	Die experimentelle Pädagogik (after vol. 5, Zeitschrift für experimentelle Pädagogik.)
I. M. S. H.	International Magazine of School Hygiene.
Iowa S.	University of Iowa Studies in Psychology.
J. E. P.	Journal of Educational Psychology.
P. A.	Psychologische Arbeiten.
P. B.	Psychological Bulletin.
P. R.	Psychological Review.
P. R. M. S.	Psychological Review Monograph Supplement.
Ph. S.	Philosophische Studien.
Pd. S.	Pedagogical Seminary.
S. Z.	Schiller-Ziehen, Sammlung von Abhandlungen aus dem Gebiet der pädagogischen Psychologie und Physiologie.
U. S.	Report United States Commissioner of Education.
Yale S.	Studies from the Yale Psychological Laboratory.
Z. A. P.	Zeitschrift für angewandte Psychologie.
Z. P.	Zeitschrift für Psychologie und Physiologie der Sinnesorgane.
Z. P. P.	Zeitschrift für pädagogische Psychologie.
Z. S.	Zeitschrift für Schulgesundheitspflege.

INTRODUCTORY

CHAPTER I

The Nature and Purpose of Mental Tests[1]

When we speak of a mental test we have in mind the experimental determination of some phase of mental capacity, the scientific measurement of some mental trait.

The mental test in some respects resembles, in some respects differs from the typical experiment of the psychological laboratory. Like this latter, the test is superior to the casual observation of everyday life because it is purposeful and methodical: it thus possesses all the merits common to experimental investigation at large, viz: the control of conditions (including the elimination of disturbing, and the systematic isolation of contributory factors), the possibility of repetition, and the possibility of subjecting the obtained results to quantitative treatment.

Unlike the typical experiment of the psychological laboratory, the mental test ordinarily places little or no emphasis upon introspective observation by the subject, in part because of its relatively short duration, in part because it is frequently applied to inexperienced subjects who are incapable of aught but the most elementary introspection, but more especially because it is concerned less with the qualitative examination or structural analysis of mental processes than with the quantitative determination of mental efficiency; because, in other words, it studies mental performance rather than mental content. It is also commonly simpler in form than the psychological experiment.

[1] The tests with which this volume is concerned are mainly mental tests. Since, however, the intimacy of the relation between mind and body makes it well-nigh imperative to study their interrelations, attention has been paid to the more important anthropometric measurements and to those tests of physical capacity that have most frequently been used in the search for correlations of psychical and physical ability.

The purposes for which mental tests have been developed are, of course, varied, but, roughly speaking, we may distinguish a theoretical interest on the part of laboratory psychologists, and a practical interest on the part of those who are concerned with mind at work in everyday life.

Historically, it appears that most of the tests now in use have originated in the psychological laboratory, either in the natural course of the development of experimental psychology as a system, *e. g.*, the usual tests of sensory discrimination, or as a consequence of special attempts to study mental capacity, particularly the interrelations of various mental capacities and of mental with physical capacities. It is, we think, not too much to hope that in time the application of mental tests will bear rich fruit in this field. We may hope that the skillful study of mental functions by the test-method may supply us with a satisfactory account of the nature and interrelations of mental functions, just as the typical introspective experiment has been able to furnish an account of the structural make-up of mind. If we could, to take an instance, obtain an exact science of mental functions so that we could know the unit-characters of mind as the biologist knows, or expects to know, the unit-characters of plants and animals, the study of mental inheritance would be carried appreciably forward.

Outside the laboratory an active and very natural interest in mental tests has been exhibited by those who are busy with practical problems to the solution of which the scientific study of mind may be expected to contribute. It is, naturally, the educator to whom the development of a significant and reliable system of mental tests would most appeal, since he is concerned with the development of just those capacities of mind that these tests propose to measure.

There has been, unfortunately it seems to us, a disposition in some quarters to speak as if a science of mental tests was already achieved; as if, for instance, a child's native ability could now be measured as easily as his height, as if his suggestibility or his capacity for concentration of attention could be determined as readily as his skull circumference or his breathing capacity. To make such assertions is surely misleading, for, as the study of

the tests herein embodied will show, there is, at the present time, scarcely a single mental test that can be applied unequivocally as a psychical measuring-rod. The fact is we have not agreed upon methods of procedure; we too often do not know what we are measuring; and we too seldom realize the astounding complexity, variety, and delicacy of form of our psychical nature.

Paradoxical as it may seem, these are the reasons, we believe, that render the elaboration of a scientific system of mental tests a possibility, for, if the all-too-evident lack of agreement in the results of the investigations already made is not attributable to faulty or divergent methods, or to clumsiness and ignorance,—if, in other words, the discrepancies are inherent and ultimate,—then we never can have a science of mental tests.

What we need is not new tests, though they are welcome enough, but an exhaustive investigation of a selected group of tests that have already been described or proposed. In particular, we need more than anything else, at least from the point of view of application, the establishment of norms of performance for these tests,—norms that are based upon investigations in which standard and prescribed methods of procedure have been followed in a rigid and undeviating manner.

This book is an attempt to assist in the realization of this need. It presents a program of work, rather than a final system of results.

CHAPTER II

General Rules for the Conduct of Tests

The following general rules may be laid down at the outset.

(1) The essential and fundamental principle underlying the conduct of scientific tests is the *standardization of conditions*. This does not mean that expensive apparatus or instruments of precision are always necessary, but simply that the conditions under which a test is given to one person or to one group of persons must be identically followed in giving the same test to another person or group. We cannot always make the conditions ideal, but we can at least try to keep them constant. If the conditions are varied, they must be varied intentionally and for a definite purpose.

(2) No detail in the 'setting' of a test is too trivial to be neglected. This is, of course, merely a restatement of the previous principle in another form. It is noteworthy that the lack of accordance between the results obtained by different investigators in the use of what is ostensibly the same test almost invariably turns out to be due to seemingly trivial variations in the method of administering the test.

In particular, attention may be called here to such matters as the time of day at which the experiment is made, the nature of the instructions that precede the test, the emotional attitude of the participants toward the investigation, their ability exactly to comprehend what is wanted of them (of which more hereafter) and their willingness to do their best throughout the test. It is well to write out the preliminary instructions and to memorize them, after first making a trial in order to see if they are perfectly intelligible. Thus, for instance, to say to one class of school children: "Cross out all the *e*'s on this paper while I take your time with a watch," and to another class: "Cross out all the *e*'s on this paper as fast as you can" may mean the same thing to

the experimenter, but it will not bring the same results from the classes under investigation, because in the second case the idea of fast work has been more strongly emphasized.

(3) No test should be undertaken by the examiner, *E*, until he is perfectly familiar with its nature, its purpose and its administration. Especially if it involves the use of apparatus, he should familiarize himself with the manipulations until they become automatic.

(4) No test should be undertaken until the subject, *S*, is perfectly clear as to what is required of him. Since most mental tests are of an unfamiliar character, something beside explicit instructions, however clearly put, is needed to enable the average *S* to undertake the test under proper conditions. Ordinarily, a brief period (say 1 to 5 minutes) of fore-exercise is needed to remove timidity, excitement or misunderstanding. If this preliminary exercise is properly arranged (especially by being based upon material not used in the test proper, and by being of the same length and character for all *S*'s), it does not introduce a serious practise error, while it does decidedly facilitate the test. In some cases, however, as, for instance, when the facility of adaptation to the test-conditions is itself an object of investigation, the fore-exercise should be omitted.

(5) *E* should be on the look-out for external signs of the way in which *S* responds to the test, *i. e.*, for indications of readiness, of quick comprehension, of a competitive spirit, or of *ennui*, fatigue, distraction, shift of attention, trickiness or deceit. The record-blanks should have a space for the recording of remarks of this nature. When tests are conducted individually it is surprising how much can be gleaned in regard to *S*'s mental traits by these indirect hints. In particular, whenever the object of the test is to examine the correlation of some physical or mental trait with *S*'s general intelligence, it is largely upon this sort of observational record that *E* must depend for his estimate of this general intelligence, even though the test be supplemented by school marks, the estimates of teachers, and similar devices.

(6) Most mental tests may be administered either to individuals or to groups. Both methods have advantages and disadvantages. The *group method* has, of course, the particular merit of economy of time; a class of 50 or 100 children may take a test

in less than a fiftieth or a hundreth of the time needed to administer the same test individually. Again, in certain comparative studies, *e. g.*, of the effects of a week's vacation upon the mental efficiency of school children, it becomes imperative that all *S*'s should take the tests at the same time. On the other hand, there are almost sure to be some *S*'s in every group that, for one reason or another, fail to follow instructions or to execute the test to the best of their ability. The individual method allows ***E*** to detect these cases, and in general, by the exercise of personal supervision, to gain, as has been noted above, valuable information concerning *S*'s attitude toward the test.

(7) One phase of the group *vs.* individual procedure demands special attention, viz: the problem of time-control. In many, if not in most tests, efficiency is measured, at least in part, by the rate at which the assigned work is performed. Now, in theory, rate or speed might be measured either by the amount of work performed within a given time or by the time taken to perform a given amount of work, in other words, by a *time-limit method* or by a *work-limit method*. In practise, however, we often find it difficult to arrange the material of a test in such a way as to make the task of equal objective difficulty at every portion of the test, and, even when this is possible, subjective variations may appear because of the fact that different *S*'s accomplish different amounts of work. There is no doubt, therefore, that the work-limit method is to be preferred to the time-limit method: it is better, in other words, that every *S* should be asked to perform the same work and to measure his efficiency in terms of elapsed time than to require every *S* to work for the same time and to measure his efficiency in 'ground covered.' But the time-limit method is compulsory in all tests of this order undertaken by groups.

(8) This leads naturally to a consideration of other difficulties that arise in scoring individual performance. Special difficulties are considered later in the discussion of the tests in which they appear, while the methods of handling measurements in general are treated in the following chapter. Attention is called here, however, to one fundamental problem, viz: *the relation of quantity of work to quality of work.* These two factors appear in nearly

every test of mental efficiency, and the question arises: shall efficiency be measured in terms of quality, excellence, delicacy or accuracy of work, or shall it be measured in terms of quantity, rate, or speed of work? To this question no general answer can be given. Roughly speaking, quantity and quality of work probably tend, at least for a given *S*, to be inversely related. Whenever this relation can be demonstrated, it is theoretically, and often, indeed, actually possible to convert the two measures into a single index of 'net efficiency,'—an index that is much desired for the study of general comparative relations. In other instances it has been proposed[1] so to adjust the conditions of the test as to throw the emphasis so strongly upon quantity or upon quality of performance that the unemphasized factor may be neglected. In yet other instances, it seems necessary to keep both an index of quantity and an index of quality, and to make reference to both in subsequent comparative study.

(9) In the application of any test, it is usual first to secure certain *preliminary data* concerning *S*'s personal history. Thus, in the experimental study of school children, *E* will find it advisable to record (*a*) name of the pupil in full, (*b*) sex, (*c*) date of birth, (*d*) name of school, (*e*) grade, (*f*) date, (*g*) hour. Other items, less uniformly recorded, but often of interest, are the following: (*h*) general health, (*i*) color of eyes and hair, (*j*) right or left-handedness, (*k*) name of teacher, (*l*) names and address of parents, (*m*) nationality of parents, (*n*) date of birth of parents, (*o*) occupation of parents, (*p*) number of children in family and their sex, (*q*) number of pupil in children of his family, (*r*) medical history of the pupil and his family, (*s*) obvious developmental defects or physical peculiarities, (*t*) details of personal habits, such as sleeping, eating, drinking, smoking, exercise, work, etc., (*u*) conduct in school, (*v*) proficiency in school work.

In *recording age* it is best to note the exact date of *S*'s birth. Unfortunately, direct comparison of the results of different investigators has at times been rendered difficult on account of disparity in the method of recording age. Thus, in arranging statis-

[1]For an illustration of both of these methods for obtaining a single index, see the Cancellation Test.

tical tables, a boy 9 years and 7 months old would by some be classed in the group of 9-year olds, by others in the group of 10-year olds, as being nearer 10 than 9. A third method, which has the advantage of being clear to the reader and not confusing to *E*, is to put all *S*'s at or past a given birthday into a single group, the age of which is specified as that birthday, plus a half-year, *e. g.*, all *S*'s between their 9th and 10th birthday comprise the 9.5 year-old group, since their average age tends, of course, to approximate 9.5 years.

CHAPTER III

The Treatment of Measures

The immediate results of the application of mental and physical tests are very apt to be obscure or unintelligible until they have been ordered and systematized by proper statistical treatment. It is the purpose of the present chapter to explain the most common methods by which this systematization is accomplished.[1]

A. MEASURES OF GENERAL TENDENCY

In many cases it is unnecessary, if not impossible, to keep in view the individual measurements of an extended series. We naturally seek to condense these values into a single representative value. Any single measure that affords us such a summary of a series of measurements may be termed a 'representative measure' or a 'measure of general tendency.' There are three such measures in common use,—the average or mean, the median, and the mode.

1. *The Mean.*

(*a*) The *ordinary arithmetical mean* (M), more often termed the *average* in psychological measurements, is computed by dividing the sum of the several measurements or magnitudes (m) by their number (n).

Hence:

$$M = \frac{\Sigma m}{n}. \tag{1}$$

[1] The reader will find more extended discussions of measurement methods in the following: Galton, Thorndike, Titchener, Sanford, Wissler, Spearman, C. B. Davenport, E. Davenport, Merriman, and Elderton (see the end of this chapter for exact references). Technical papers upon correlation formulas by Pearson, Yule, and others will be found in various numbers of *Biometrika*, the *Proc. of the Royal Soc.* of London, and in the *Phil. Transactions* of the same body.

TABLE I

Strength of Grip, in Hectograms, 50 boys (Whipple)

ORDER	RIGHT HAND				LEFT HAND				RANK COMPARISONS			
	NO.	STANDING	*d*	*d*²	NO.	STANDING	*d*	*d*²	G	D	D²	*xy*
1	30	158	−125	15625	30	138	−135	18225	—	0	0	16740
2	17	175	−108	11664	17	163	−110	12100	—	0	0	11880
3	52	193	− 90	8100	1	175	− 98	9604	—	6	36	7470
4	39	197	− 86	7396	48	180	− 93	8649	—	2	4	7568
5	10	197	− 86	7396	7	180	− 93	8649	—	9	81	5246
6	43	200	− 83	6889	39	185	− 88	7744	—	2	4	6889
7	1	205	− 78	6084	16	190	− 83	6889	4	4	16	7644
8	40	206	− 77	5929	43	190	− 83	6889	—	4	16	5236
9	3	208	− 75	5625	52	190	− 83	6889	—	1	1	5550
10	7	210	− 73	5329	3	199	− 74	5476	5	5	25	6789
11	6	210	− 73	5329	6	200	− 73	5329	—	0	0	5329
12	48	220	− 63	3969	40	205	− 68	4624	8	8	64	5859
13	42	225	− 58	3364	15	210	− 63	3969	—	6	36	2204
14	2	225	− 58	3364	10	212	− 61	3721	—	2	4	2842
15	19	225	− 58	3364	19	215	− 58	3364	—	0	0	3364
16	37	226	− 57	3249	2	224	− 49	2401	—	10	100	741
17	15	235	− 48	2304	50	235	− 38	1444	4	4	16	3024
18	8	244	− 39	1521	8	235	− 38	1444	—	0	0	1482
19	14	244	− 39	1521	42	235	− 38	1444	—	6	36	507
20	51	245	− 38	1444	23	242	− 31	961	—	1	1	1102
21	50	248	− 35	1156	51	244	− 29	841	4	4	16	1330
22	41	262	− 21	441	45	245	− 28	784	—	2	4	273
23	25	262	− 21	441	9	253	− 20	400	—	14	196	−735
24	23	267	− 16	256	41	260	− 13	169	4	4	16	496
25	12	269	− 14	196	14	260	− 13	169	—	5	25	− 98
26	44	270	− 13	169	37	260	− 13	169	—	1	1	78
27	29	273	− 10	100	44	267	− 6	36	—	2	4	− 20
28	9	280	− 3	9	34	270	− 3	9	5	5	25	60
29	45	290	+ 7	49	29	275	+ 2	4	7	7	49	−196
30	36	294	11	121	12	280	7	49	—	10	100	594
31	11	296	13	169	32	282	9	81	—	3	9	312
32	28	301	18	324	20	290	17	289	—	4	16	576
33	32	310	27	729	13	290	17	289	2	2	4	243
34	31	313	30	900	11	297	24	576	—	1	1	810
35	20	315	32	1024	31	300	27	729	3	3	9	544
36	34	320	37	1369	28	305	32	1024	8	8	64	− 111
37	24	323	40	1600	25	308	35	1225	—	4	16	2520
38	16	325	42	1764	26	315	42	1764	31	31	961	−3486
39	35	330	47	2209	38	325	52	2704	—	3	9	3854
40	13	346	63	3969	36	327	54	2916	7	7	49	1071

TABLE 1 (Continued)

ORDER	RIGHT HAND NO.	STAND-ING	d	d^2	LEFT HAND NO.	STAND-ING	d	d^2	RANK COMPARISONS G	D	D^2	xy
41	38	348	65	4225	24	336	63	3969	2	2	4	3380
42	46	350	67	4489	35	355	82	6724	—	3	9	7035
43	33	353	70	4900	49	362	89	7921	—	1	1	7140
44	26	375	92	8649	33	375	102	10404	6	6	36	3864
45	21	375	92	8649	46	378	105	11025	—	2	4	12236
46	18	403	120	14400	18	400	127	16129	—	0	0	15240
47	27	430	143	20449	21	406	133	17689	—	2	4	25311
48	49	440	153	23409	5	443	170	28900	5	5	25	13617
49	5	440	153	23409	27	450	177	31329	1	1	1	26010
50	22	508	221	48841	22	490	217	47089	—	0	0	47957
Sums		14164	3088	287884		13651	3166	315223	106		2098	277371
Aver...		283	61.4	5757.7		273	63.3	6304.4				

r = [illegible] = 6

TABLE 2

Values Derived from the Data of Table 1

MEASURE	FORMULA	VALUE Right Hand	Left Hand
Mean	1	283.0	273.0
Median	3	269.5	260.0
A. D.	4	61.4	63.3
S. D.	5	75.8	79.0
S. D.	6	76.5	80.2
S. D.	7	76.9	79.3
P. E.	10	51.1	53.3
P. E.	11	51.6	54.0
P. E.	12	50.7	52.0
P. E.	13	51.9	53.5
C.	16	0.27	0.29

The mean is the most familiar measure of general tendency, and it is the most precise, because it is affected by all measurements in proportion to their size. It has, however, some disadvantages: its computation requires more labor than that of the median or the mode,[4] and, as will be shown later, it may fail after all to afford a truly representative value. Examples of arithmetical means are scarcely needed, but may be found in Tables 1 and 2, and elsewhere.

(*b*) The *weighted arithmetical mean* is serviceable as a short cut in dealing with a large number of measures. Its use may be made evident by the following hypothetical case. Suppose it were desired to ascertain the average height of 1000 12-year old boys. By the ordinary method we should be obliged to record each measure exactly (say, within 1 mm.) and to add the entire 1000 measurements. To utilize the weighted arithmetical mean, we divide the range of height into a limited number of groups of, let us say, 2 cm., and record simply the number of cases that fall into each group, *i. e.*, the frequency of each group. Thus in Table 3, there are in the 6th group 88 measurements lying between the limits 135 and 137 cm. The weighted mean can now be found very simply by multiplying the value or magnitude representing each group by the corresponding frequency (1×126, 5×128, etc.) and dividing the sum of the products by the sum of the frequencies (1000).

The formula for the weighted arithmetical mean is therefore:

$$M = \frac{\Sigma\,(m \cdot f)}{\Sigma f},$$

or:

$$M = \frac{\Sigma\,(m \cdot f)}{n}. \tag{2}$$

It is clear that this weighted mean approaches the ordinary mean in accuracy in proportion as the number of classificatory groups is increased.

[4] The computation of M may be greatly lessened by assuming a convenient approximate value, and correcting subsequently to the true value. For illustrations, see Davenport (4, p. 429) or Thorndike (21, p. 71).

TABLE 3

Distribution of the Heights of 12-Year Old Boys (Hypothetical)

Centimeters	126	128	130	132	134	136	138	140	142	144	146	148	150	152	154	156	158	160
Deviation..	−16	−14	−12	−10	−8	−6	−4	−2	0	2	4	6	8	10	12	14	16	18
Frequency.	1	5	14	24	39	58	96	120	150	142	123	88	63	36	23	12	5	1

Weighted mean = 142.9 Median = 143.4 Mode = 142.

2. The Median.

The ***median*** or ***central value*** is, literally, the middlemost of a group of measurements arranged singly in ascending or descending order, or the measure above and below which lie an equal number of individual measurements. It is expressed, therefore, by the formula:

$$\text{median} = \text{the } \frac{n+1}{2} \text{ measurement} \qquad (3)$$

In practise the median may or may not coincide with some actual measurement; more often than not it is an interpolated value. To compute its value we must first arrange the measures serially (if any magnitude is repeated two or more times, the number of such repetitions must, of course, be indicated). To find the middlemost measure when interpolation is needed, we may proceed by a simple method which may be illustrated by reference to Table 3. Here the first 8 groups (126 to 140 cm.) represent 357 measurements; since there are 1000 measurements, the middlemost measurement lies in the ninth group of 150 cases: the value desired is, therefore, the theoretical value between the 143d and the 144th measurement in this group. We have, therefore, to take $\frac{143.5}{150}$ of the range of magnitude covered by the 9th group (2 cm.), and add this to the lower limiting value of the group, 141.5 cm., so that we obtain for the median the value, 141.5 + 1.9, or 143.4.

The great merit of the median is the ease with which it can be determined: in short series it is not even necessary to arrange

the measures serially, as one-half of the measurements may be checked off by inspection. Its primary disadvantage is that it gives little weight to extreme deviations and may fail entirely to represent the type, yet, in many psychological observations, it is precisely these extreme deviations which are most suspicious, so that this tendency of the median to lessen the significance of extreme measures may prove a positive advantage.

In general, the longer the series or the more homogeneous the values, the more nearly does the median approximate the mean.

3. *The Mode.*

If a number of measurements are distributed in ascending or descending order, a *mode* is a measure that appears more frequently than do measures just above or below it in the series. There may be several modes in a distribution, though usually there is but one, and we may therefore define the mode as the commonest single value, or the commonest condition.

Many statistical arrays find a better representative value in the mode than in the average. Thus, when we speak of the "average American citizen," we really have in mind the typical citizen, the one most frequently met with. To borrow an illustration from Rietz (4, p. 684): "If a community has 10 millionaires, but all the other citizens are in poverty, an arithmetical average might give the impression that the people of the community are in good financial condition, while really the 'aveiage citizen' is in poverty." The primary use of the mode is therefore, *to characterize a type.*

Strictly speaking, we may have an empirical mode, as indicated in a given array and a theoretical mode, which would be the most frequent condition in a theoretical distribution. The latter is difficult to compute and not often employed. If an array is very irregular, there is, in strictness, no mode or type at all, or at least the indicated mode has little significance.

In Table 3, it is clear that the mode is 142 cm., because this measure appears 150 times, and no other measure is as frequent.

B. MEASURES OF VARIABILITY

It is a common fallacy to rest content with the statement of the general tendency of measurements. Even in supposedly accurate and scientific determinations, we may find the quantitative expression limited to averages, *e.g.*, "the mean temperature for September," "the average weight of 12-year old boys," etc. But it is evident that the average gives no indication of the distribution of the individual measures from which it is obtained, no indication of the extent to which these measures vary or deviate from the average, no information as to how homogeneous is the material that the average represents. The September temperature may have been seasonable and equable or there may have been some days of frost and some days of sweltering heat. Again, if five individuals weigh 80, 65, 60, 40, and 55 kg., respectively, and five others 62, 59, 60, 51, and 58 kg., respectively, then the mean weight of either group is 60 kg., but one group is distributed very closely around the mean, whereas the other group exhibits such marked deviations from it that M (or any other general tendency measure) has little or no significance as a representative value.

From this it follows that we need not only measures of general tendency, but also measures of the variability or tendency to deviation of measurements, and that these latter are of well-nigh equal importance.

There are three common measures of variability,[1]—the average deviation, the standard deviation, and the probable error.[2]

1. *The Average Deviation* (*Mean Variation*)

To find the average deviation we must first find the mean, M, (or median or mode); second, substract each individual measure m, algebraically from M, which gives a series of deviations, d;

[1] Besides these measures, range of variability is sometimes indicated roughly by stating the maximal and minimal measurements, in conjunction with M. This gives us, at least, information as to the extremes of deviation.

[2] It is well to avoid confusion here at the outset. The average deviation (A. D.), as used by the statisticians, is identical with the mean variation (m. v.) of experimental psychology. The standard deviation (σ) is called the average error by Sanford, the mean error by Merriman, and the error of mean square by others.

third, find the average of these deviations, *i. e.*, the mean of the variations, by summating without reference to sign and dividing by the number of cases.
Hence:

$$A.D. \text{ or } m.v. = \frac{(M - m_1) + (M - m_2) + \ldots\ldots(M - m_n)}{n},$$

or:

$$A.D. = \frac{d_1 + d_2 + \ldots d_n}{n},$$

or:

$$A.D. = \frac{\Sigma\, d}{n}. \tag{4}$$

Reference to Table 1 will render this process clear: there the average right hand grip is 283; the weakest boy has a standing of 158, hence he deviates 125 units from the average; the first 28 boys rank below average and therefore exhibit minus deviations, the rest are above average and exhibit plus deviations; all these deviations are added without regard to sign and their sum, 3088, is divided by the number of cases, 50, yielding a mean variation of 61.4 hectograms. If the median were selected as the representative value, the variability would, of course, be computed similarly with a new series of *d*'s.

2. The Standard Deviation (Error of Mean Square)

This measure of variability is preferred by many experimenters and is practically the only one employed by statisticians, as it is thought to be more accurate than the average deviation, but it is much more laborious to compute. It is the square root of the average of the squares of the individual deviations

$$S.\,D.\text{, or } \sigma = \sqrt{\frac{d_1^2 + d_2^2 + d_3^2 + \ldots d_n^2}{n}},$$

or:

$$\sigma = \sqrt{\frac{\Sigma\,(d)^2}{n}}. \tag{5}$$

If n is small, the formula is often modified by writing $n-1$ in place of n:[1]
Hence:

$$\sigma = \sqrt{\frac{\Sigma\,(d)^2}{n-1}}. \tag{6}$$

The application of Formula 5 is illustrated in Table 1, 5th and 9th columns, where the squares of the individual deviations are shown in detail. The sum of these squares foi the right-hand grip is 287,884. This is divided by 50, giving 5757.7, the square root of which is 75.8, the σ desired.

The *S. D.* of a given series is somewhat larger than its *A. D.* Theoretically, and practically if the distribution be symmetrical and the observations sufficiently numerous, the relation is constant at

$$\sigma = 1.2533\ A.\ D. \tag{7}$$

Conversely,

$$A.\,D. = 0.7979\,\sigma. \tag{8}$$

As shown in Table 2, the *S. D.* computed by Formula 7 is closely similar to that computed by Formula 6.

3. *The Probable Error*

The probable error of a single measure (*P. E.*) is a measure of the limits above and below M (or other representative measure) that will include one-half of the individual measures; in other words, it is a value such that the number of measures that exceed it is the same as the number of measures that fail to reach it.[2]

[1] For the reasons for this substitution, consult Merriman (p. 71). It is evident that the effect of the substitution becomes progressively less as n increases: as will be seen in Table 2, the difference between Formula 5 and Formula 6 is practically negligible when $n = 50$.

[2] The term 'probable error' is often a source of confusion to those unfamiliar with its use in mathematics. The magnitude in question is not, of course, the *most* probable error, neither is it, from our point of view, an 'error' at all. For a descriptive term, we might call the probable error the median deviation since it is that deviation that is found midway from the representative value in either direction.

The *P. E.* is approximately two-thirds the *S. D.*, or more exactly.

$$P.\,E. = 0.6745\,\sigma. \tag{9}$$

By reference to Formula 5 this becomes:

$$P.E. = 0.6745\sqrt{\frac{\Sigma\,(d)^2}{n}}, \tag{10}$$

or, for a small number of cases (Formula 6):

$$P.E. = 0.6745\sqrt{\frac{\Sigma\,(d)^2}{n-1}}. \tag{11}$$

In practise we may find the *P. E.* approximately, if the distribution be assumed to be normal (see under *D*, below), by counting off one-fourth of the cases from either end of a series of measurements, and halving the difference between the two values thus found.

$$P.E. = \frac{m_{\frac{3}{4}n} - m_{\frac{1}{4}n}}{2}. \tag{12}$$

Thus in Table 1, these limits lie at the 12th and a half and the 37th and a half measurements, and have, for the right-hand grip the values 222.5 and 324, respectively; hence, $P.\ E. = 324 - 222.5 \div 2 = 50.7$,—a value that is approximately the same as the values of *P. E.* computed by Formulas 10, 11, and 13 (Table 2). By Formula 11, $P.\ E. = 0.6745 \times 76.5 = 51.6$. Corresponding values are given in Table 2 for the left-hand grip as distributed in Table 1. Still other values might be computed on the basis of the median instead of the mean.

By combination of Formulas 7, 8, and 9, we may obtain for a normal distribution:

$$P.\,E. = 0.8453\ A.\,D. \tag{13}$$

$$S.\,D. = 1.4825\ P.\,E. \tag{14}$$

$$A.\,D. = 1.1843\ P.\,E. \tag{15}$$

The first of these is illustrated in Table 2.

4. The Coefficient of Variability

If it is desired to compare the variability of one series of measurements with that of another, it will be found that, as a rule, their respective measures of variability cannot be compared directly, because they are based upon different units or at least upon different measures of general tendency, but the relations of the two measures of variability to their respective measures of general tendency can be directly compared. In other words, we can compute two coefficients of variability (C) by dividing in each series a measure of variability by a representative measure, *i. e.*, either *S. D.*, *A. D.*, or *P. E.*, may be divided by either mean, median, or mode. Unless otherwise specified, it may be assumed that *S. D.* is divided by M.

Hence:

$$C = \frac{\sigma}{M} \tag{16}$$

Thus, in Table 1, for strength of right hand, $C = 76.5 \div 283 = .27$ and for strength of left hand, $C = 80.2 \div 273 = .29$, hence the latter series is slightly more variable.

C. THE GRAPHIC REPRESENTATION OF MEASUREMENTS

A series of measurements, as we have seen, can be expressed adequately by a single representative value only when that value is accompanied by some measure of variability. Even these two values may fail to express the series completely, since they are, after all, only symbols for the convenient summarizing of general tendency and variability, whereas a complete numerical expression of a series of measures would imply the tabulation of all the data of the series. Such a tabulation is for the most part impracticable, or at least of little significance, because of the difficulty of grasping the nature of the series by the inspection of a mass of figures.

The use of the graphic method, however, supplies a most serviceable and effective means of showing at a glance all of the important features in the distribution of a series of measurements and likewise of relations between series of measurements.

1. *The Plotting of Frequencies or Graphs of Distribution*

The most usual form of graph for illustrating the distribution of a series of measurements is constructed as follows:

Draw two lines, OY and OX (Fig. 1) in the form of coördinate axes, *i. e.*, with OY perpendicular to OX. Upon the horizontal, or x-axis, lay off convenient intervals corresponding to the units of measurement of the series to be plotted; upon the vertical, or y-axis, lay off intervals corresponding to the frequencies of the series.

The choice of the scale units is largely arbitrary. The intervals of the two axes need not be the same, nor need different graphs, save for purposes of direct comparison, be plotted to the same scale. In general, a scale should be selected that will bring the surface easily into view as a whole and that will render conspicuous the features that are under consideration. Thus, if one is studying rate of increase or decrease, a scale should be selected that affords a fairly steep curve in order to emphasize its rise and fall. 'Squared' or cross-section paper (usually laid off by mm. on sheets 15 x 20 cm.) may be purchased for curve-plotting, and will be found invaluable for this work.

In illustration, the numerical table of frequencies above (Table 3) is turned into a surface of frequency upon the axes just mentioned (Fig. 1). We mark off on the x-axis, it will be seen, 18 equal intervals corresponding to the range of dimensions, 126, 128, . . . 160 cm. Upon the y-axis we mark off equidistant intervals for the range of frequencies from 1 to 150. We next locate the series of 18 points. The first point lies vertically above the 126 cm. mark at a distance equal to 1 of the vertical units; the second lies vertically above the 128 cm. mark at a distance equal to 5 vertical units, etc. By joining the 18 points thus located, the resulting line evidently gives in a single visual impression the distribution that was expressed numerically in Table 3. Any point in this line is fixed by stating its abscissa or distance from the y-axis, and its ordinate, or distance from the x-axis.[1]

Now it would have been equally feasible to have considered the values in Table 3 in terms of their deviation from the mean, median or mode, and with little or no change in the curve. Take, for simplic-

[1] The 'curve' is sometimes so drawn as to form the tops of a series of columns erected at the intervals on the base-line, instead of by joining the single points as here described. See, for illustrations, Thorndike (21, p. 48, or 20, p. 15).

ity, the mode, 142 cm., as the representative value. Erect an ordinate of the value of 150 at a point **M** on the x-axis (Fig. 1); intervals to the right of this ordinate may now represent positive deviations (+ 2, + 4, + 6, etc.) while those to the left represent negative deviations (-2, -4, -6, etc.), as indicated in Table 3. It thus becomes possible to represent negative values graphically.

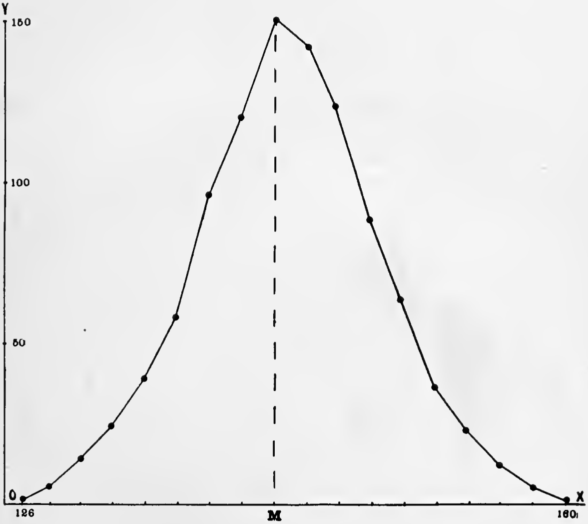


FIG. 1. GRAPHIC REPRESENTATION OF THE DISTRIBUTION OF TABLE 3.

2. *The 'Smoothing' of Distributions*

Ordinary measurements are subject to numerous disturbing factors; our units of measurement are often coarse; our opportunities for securing data are always restricted; variable factors of one sort or another obtrude themselves,— and these disturbances produce irregularities in the resultant data. The obtained distribution, in other words, does not coincide with the true distri-

bution, *i. e.*, with the distribution that would theoretically appear under ideal conditions. Thus, in Table 3, chance may have led to the measurement among the 1,000 cases studied, of more boys of a certain height, say 144 cm., than we should ordinarily have encountered in measuring 1,000 pupils taken at random. Or, to take an instance of a striking artificial distortion, in the census returns, people who are 39 or 41 years of age show a tendency to report their age as 40, so that the age of 40 has an unnaturally large frequency.

Minor deviations from the theoretically expected distribution may be counteracted if we are constructing a frequency graph by 'smoothing' the curve, *i. e.*, by drawing the connecting line in the form of a true curve rather than a broken straight line: such a curve will pass in the neighborhood of the several points which have been located by the numerical data, but will not necessarily pass exactly through these points. The result is a graph that shows how the data would presumably have been distributed if the factors which produced the distortions and irregularities were eliminated.

TABLE 4

The Numerical Smoothing of the Distribution of Table 3

Centimeters	126	128	130	132	134	136	138	140	142	144	146	148	150	152	154	156	158	160
Original...	1	5	14	24	39	58	96	120	150	142	123	88	63	36	23	12	5	1
Smoothed .	2	6	14	26	40	64	91	122	137	138	108	91	62	41	24	13	6	2

In tabular work these deviations may be counteracted by a simple arithmetical process. Replace each frequency except the two extreme ones by the mean (to the nearest integer) of the given frequency and the one just below it and the one just above it; replace the two extreme frequencies by the mean (to the nearest integer) of the given frequency taken two times and the adjacent frequency taken once. If necessary, a second smoothing may be made of the values obtained by the first smoothing.

The values of Table 3 do not exhibit marked irregularities as is evident from their graphic distribution in Fig. 1: the process of smoothing may, however, be illustrated by the treatment in Table 4.

D. NORMAL AND OTHER TYPES OF DISTRIBUTION: THE PROBABILITY SURFACE AND ITS APPLICATIONS

1. The Normal Frequency Surface

Assume that errors of observation have been eliminated and that a large number of measurements of some psychological trait or capacity have been secured: experience has shown, and theoretical considerations likewise indicate, that as a rule these measurements will distribute themselves in the form of a symmetrical bell-shaped curve, variously known as the probability curve, the curve of error, Gauss' curve, or the normal frequency surface,—the salient characteristics of which are a maximal frequency at M with a series of positive and negative d's from M that are symmetrically disposed on either side of it and whose frequency decreases progressively as their size increases.

Such a distribution implies the operation in the conditions that underlie the feature or trait under measurement, of an indefinitely large number of individual factors, each of which is equally likely to be present and effective. When, however, there are limiting or restricting conditions, or when one or more factors are present oftener than mere chance would allow, the resultant distribution will tend to depart from the normal type. Thus, the chances of death at different ages are not distributed according to the normal curve, but are higher in infancy and old age than in youth and middle age. The mental ability of college students is not likely to be distributed like that of the non-college population of the same age on account of the selective influence of entrance requirements.[1] In general, distributions that do not conform to the normal type are termed 'skewed' distributions, and may demand special mathematical treatment.

2. Relation of the Normal Curve to S. D. and P. E.

The normal surface of frequency has interest still further because in it the significance of *P. E.* and of *S. D.* becomes clear. In fact, the latter bears to the curve a relation like that of a radius

[1] On the application of the normal curve to the grading of college students, see M. Meyer (11).

to its circle. If *S. D.* is small, the measurements are relatively homogeneous and the curve is steep and compact (right-hand curve in Fig. 2), whereas, if *S. D.* is large, the curve is broad and of easy slope (left hand curve in Fig. 2). If *M* and *S. D.* are known, the entire curve for a normal distribution is known. If the distribution is not of the normal form, the *S. D.* still remains a good measure of its variability, though not completely descriptive of the entire distribution.[1]

The geometrical explanation of the *P. E.* is simple. In Fig. 2 we draw the ordinates *ab* and *cd* equidistant from *OY* and at such a distance that the area *abYcd* is equal to the remainder of the total area under the curve: then the abscissa *Oa* or *Od* represents the value of *P. E.*, *i. e.*, a deviation from the mean that will include one-half the total deviations.

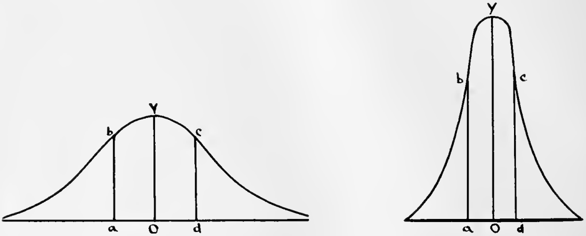


FIG. 2. TYPICAL CURVES OF NORMAL DISTRIBUTION.

3. *P. E. of M and of Other Measures*

Since our opportunities for securing data are limited, it follows that even averages may fail to be absolutely exact measures of the general tendency of the trait under measurement. To revert to the hypothetical data of Table 3, we were there able to obtain an *M*, 142.9 cm., of the height of 12-year old boys: it must be evident that if we could have measured a million boys we should feel

[1] This mathematical relation of *S. D.* to the probability curve, together with the possibility, as is shown later, of determining many other features of the distribution from the relation of *S. D.* and *M*, is one of the principal reasons why *S. D.* is preferred by many to the more-easily calculated *A. D.*

surer that the M then obtained was the true one, or that if we had measured only ten boys of that age we should not have felt at all sure that the average thus obtained was truly representative of the height of 12-year old boys. We need, therefore, a measure of the reliability of M, so that we may have some idea as to how far the actually obtained M is likely to differ from the ideal or true M, or, reversely, how many measurements we need to secure an M that will have any desired or assigned degree of reliability.

The most common measure of the *reliability of M* is afforded by its $P.\ E.$, for there can be a $P.\ E.$ of M as well as a $P.\ E.$ of a single measurement or observation. To illustrate, suppose we did measure 1,000,000 boys in 1000 groups of 1000 measurements each; if we then averaged each group we should obtain 1000 M's, each representing the central tendency of a group chosen by random sampling: we should then expect these 1000 M's to be closely similar, but not identical, and we could distribute them like a series of individual measures and determine the $P.\ E.$ of this distribution. In practise, the $P.\ E.$ of M is found by a formula that takes into consideration the variability of the distribution which M represents and the number of cases on which it is based. This formula is

$$P.\,E._M = \frac{0.6745\,\sigma}{\sqrt{n}}. \tag{17}$$

That is, the $P.\ E.$ of M is found by dividing the $P.\ E.$ of a single measurement (Formula 9) by the square root of the number of measurements.

To bring this formula into relation with $A.\ D.$, we may use the approximate formula

$$P.\,E._M = \frac{0.8453}{\sqrt{n-1}}\,A.\,D. \tag{18}$$

The consistency of a series of measurements may also be indicated by stating the degree of probability that will attach to the appearance of an 'error' or deviation or residual, as it is often termed, of a magnitude equal to any assigned multiple of $P.\ E.$ By definition, a deviation of the magnitude of $P.\ E.$ is one as likely to be exceeded as not; in other words, the chances are even, or

one to one, that it is exceeded. The probability of the occurrence of a deviation several times as large as *P. E.* is, however, very much smaller, as will be seen in the following comparisons between *P*, the theoretical probability and $X \div P.\ E.$, multiples of *P. E.*, from 1 to 5.[1]

$x \div P.E.$	$P.$
1	$1 \div 1.0$
2	$1 \div 5.6$
3	$1 \div 23.2$
4	$1 \div 143.3$
5	$1 \div 1342.2$

Besides the *P. E.* of *m* and of *M*, we may determine, by appropriate formulas, the *P. E. of measures of variability*: two of these are given below; the application to measures of relationship will be discussed later.

The *P. E.* of the *S. D.* is found by the formula

$$P.\,E. = \frac{P.\,E._M}{\sqrt{2}}, \tag{19}$$

which, by reference to Formula 18, becomes

$$P.\,E. = \frac{0.6745}{\sqrt{2\,n}}. \tag{20}$$

[1] These values are computed by reference to standard tables of values of the probability integral corresponding to various multiples of *P. E.* A condensed table of this sort is published by Thorndike (21, p. 149). The values given above were derived for the author by Prof. G. C. Comstock of the University of Wisconsin from Oppolzer's 10-figure table of the Gamma Integral, and are correct to the first place of decimals given. To illustrate from Thorndike's condensed table; the total area of the probability surface being 1000, the total area representing deviation in either direction is 500. From the table we see that a deviation or residual equal to 3 P. E. occurs in such a manner that 479 of the 500 cases are included between it and the average or median, and hence it is exceeded by 21 of 500 cases, or by 1 case in 23.8, approximately; 23.2 when more accurate integral tables are used.

From such a series of values the consistency of the determination may be stated in various ways. For example, if a correlation of .50 was accompanied by a *P. E.* of .10, it might be said that the chances would be but 1 in more than 1300 times that such a correlation would occur by mere chance.

The $P. E.$ of the coefficient of variability may be found approximately, if C is not greater than 10 per cent, by the formula

$$P. E._C = \frac{0.6745\, C}{\sqrt{2\, n}}, \tag{21}$$

but more accurately, for any value of C, by the formula

$$P. E._C = \frac{0.6745\, C}{\sqrt{2 n}} \left[1 + 2 \left(\frac{C}{100} \right)^2 \right]^{\frac{1}{2}} \tag{22}$$

4. *Other Applications of the Probability Curve*

Since, when the distribution is normal, the surface of frequency is determined by M and $S.\ D.$, we may, by reference to suitable tables, ascertain (1) the frequency of any deviation, (2) the range of deviation that will include any given percentage of m's, (3) the chances that the true M will differ from the obtained by any given amount, (4) the range of divergence of the true from the obtained M that corresponds to any given degree of improbability, and (5) in general, the degree of reliability, or unreliability, of the several measures of variability or relationship.[1]

E. MEASURES OF CORRELATION

1. *The Meaning of Correlation*

Physical science discovers numerous uniformities or correspondences between natural phenomena which are formulated as 'natural laws': biological science, on account of the intricacy of the factors which condition vital phenomena, can discover, for

[1] All of these calculations are made in terms of a probability integral table, which indicates for any normal surface the proportionate area of the probability surface that is represented by any given degree of deviation (measured in this case in terms of the $S.\ D.$ of the distribution). Lack of space precludes the discussion and reproduction of this table, which may be found in nearly all works on probability and statistics, *e. g.*, Merriman (p. 187), C. Davenport (p. 55), Scripture, *New Psychology*, p. 475, Thorndike (21, p. 148; 20. p. 168). Examples of the calculations mentioned may be studied in Thorndike (21).

the most part, only *tendencies* to uniformity or tendencies to correspondence. Such a tendency of two or more traits or capacities to vary together is termed a correlation. Thus height and weight are obviously correlated because in general tall people are heavier than short people, but, of course, this tendency to correspondence is far from absolute, like the correspondence between the distance and speed of a body falling in vacuum or between the electrical constants, voltage, amperage and resistance as expressed in Ohm's law.

Since in practically every psychological test we are searching for these tendencies toward correspondence, it is important to know how they can be measured. In not a few psychological investigations correlation has been expressed merely descriptively as 'fair,' 'large,' 'poor,' etc., and these characterizations have been derived from mere inspection of arrays of data. As a matter of fact, some of these published statements of correlation are actually wrong: correlations do not exist where they have been affirmed, or do exist where they have been denied. At the present time there is no excuse for such merely descriptive statements of correlation, since, by the use of appropriate mathematical procedure, a tendency toward correspondence may be measured and expressed by a single quantitative symbol that has as much significance and definiteness as *M*, *S. D.*, or any other statistical constant. This symbol, r, which sums up the proportionality or degree of relationship between two factors or events, is known as the *index or coefficient of correlation.*

Complete positive or direct correlation between two traits is present when the existence of the one is invariably accompanied by the existence of the other, or when increase of the one is invariably accompanied by corresponding and proportional increase of the other.

Complete negative or inverse correlation is present when two traits are mutually exclusive, or when increase in the one is invariably accompanied by a corresponding and proportional decrease in the other.

A correlation is *indifferent* or *zero* if the existence or variation of one trait is totally unrelated to that of the other.

In perfect positive correlation, r is unity or 1.00; in complete

negative correlation, r is -1.00; indifference or complete absence of correlation is 0. In actual psychological investigation, at least when functional correspondences are under investigation, we have commonly to deal with some intermediate degree of correlation, and r assumes, therefore, the form of a decimal lying between 0 and 1.00 for positive and between 0 and -1.00 for negative correspondence.[1]

2. The Computation of the Index of Correlation

(a) The 'Product-Moments' Method of Pearson

The most elaborate as well as the theoretically best possible method of computing r is the standard 'product-moments' method elaborated by Bravais, Galton, and especially by Pearson.

In referring to the product-moments as the best possible method, certain qualifications must be kept in mind. It is possible, for instance, that, as Spearman contends (17), the comparison of ranks (R-method), for reasons that will be explained later, may be more reliable and satisfactory when psychological data are under treatment.

In any case it is to be remembered that the product-moments method applies primarily to the relation of those arrays whose distribution is 'normal'—in the sense already explained. The mathematics of correlation for skewed, multimodal and other complex forms of distribution have received attention, especially at the hands of European mathematicians, but it cannot be said that their results have as yet reached a stage where they are of practical usefulness to those who have to take their measurement-methods at second-hand. (Cf. also Krueger and Spearman, pp. 53–4.)

Again, it is also possible that functional relations may exist of so simple a form as to be readily expressed either verbally or mathematically, which, nevertheless, will give a zero coefficient by the standard method. Take, for instance, the hypothetical relation that Spearman has adduced as an example of zero correlation. Suppose five persons are tested for vision and hearing with the following results (in terms of feet that the test-type is read and the sound heard):

Person	*A*	*B*	*C*	*D*	*E*
Vision, in feet	6	7	9	11	14
Hearing, in feet	6	11	12	10	8

[1] On the meaning of r, especially in procedure by the method of rank-comparison, see the discussion between Dürr and Spearman (5; 18).

Here, says Spearman (15, p. 77), "we get $r = 0$, and thus there is no correlation, direct or inverse." But, as Lehmann and Pedersen (9, pp. 16,17) have cleverly shown, if these values are plotted graphically, there is revealed a simple functional relation, viz: hearing is poor when sight is poor, reaches a maximum when sight is fairly good, and then declines when sight continues to improve. From this it may be seen that it is best to plot relations in graphic form whenever possible. In many cases, indeed, such a functional graph is more significant than any coefficient could possibly be, just as a curve of distribution is more significant than an M, even when coupled with its measure of deviation.

By the product-moments method

$$r = \frac{\Sigma\, xy}{n\ \sigma_1\, \sigma_2}, \tag{23}$$

in which the x's are the series of deviations from M in the first array and the y's the corresponding series of deviations in the second array, and in which σ_1 is the standard deviation of the first and σ_2 the standard deviation of the second array, and n the number of cases in either array.

The various steps of the computation may be illustrated by reference to Table 1[1] for grip of right and left hands, as follows:

(1) Arrange the original measurements in order of their standing or rank, as shown in Columns 1, 3, and 7. (While this is not absolutely necessary, it commonly facilitates computation, though for speedier determination of the xy values, it might be preferable to place the two arrays in the same order by individuals, *e. g.*, as shown by the numbers in Columns 2 and 6.)

(2) Compute M (or the median) for each series (283 and 273).

(3) Compute and record the individual deviations (d, columns 4 and 8) for each series, retaining the algebraic signs.

(4) Multiply the d of each individual in the first series (now termed his x) by the d for the same individual in the second array (now termed his y), and record the products, observing the algebraic signs (the xy values in Column 13), *e. g.*, boy No. 30 has for his $x, -125$ and for his $y, -135$, hence, for his xy, $-125 \times -135 = 16740$. Again, boy No. 25 has for his corresponding values, -21 and $+35$, hence for xy, -735.

[1] For a fuller illustration of correlation arithmetic, together with suggestions for shortening the work, see E. Davenport, pp. 455–471.

(5) Add the products obtained in (4) (277,371, Column 13).

(6) Compute the *S. D.* of both series (σ_2 and σ_2 in Formula 23, illustrated in the d^2 columns, 5th and 9th, and explained in Formula 5): multiply them together, and multiply their product by the number of cases ($75.8 \times 79 \times 50 = 299{,}410$).

(7) Divide the 5th by the 6th resultant for the index desired ($r = 277{,}371 \div 299{,}410 = +\ 0.93$).

The arithmetic of the Pearson method is thus simple, though somewhat tedious. The work may be materially lessened by the use of Barlow's *Tables of Squares, Cubes, Square Roots*, etc. New York, 1904, of Crelle's *Rechentafeln* (procurable through G. E. Stechert & Co., New York), which show at a glance the products of all numbers up to 1000×1000, and by the use of an adding machine.[1] Another considerable shortening may often be effected without serious disturbance by substituting Formula 7 for Formula 5 in computing the two *S. D.'s*. Thus, in our illustration, this substitution (see Table 3) gives for the denominator of the fraction: $76.9 \times 79.3 \times 50 = 304{,}908.50$, from which we find $r = 0.91$.

The *probable error of the coefficient of correlation* as obtained by the Pearson method is calculated by the formula

$$P.E._r = 0.6745\ \frac{1 - r^2}{\sqrt{n}}, \tag{24}$$

although some mathematicians prefer the formula

$$P.E._r = 0.6745\ \frac{1 - r^2}{\sqrt{n\,(1 + r^2)}}. \tag{25}$$

It is evident that the reliability of a coefficient increases with the number of cases compared and also with the magnitude of the r obtained. The actual values of $P.E._r$, as computed by Formula 24 for eleven values of r from 0 to 1 accompanying values of n from 25 to 1000, are indicated in Table 5, so that one can not only read at a glance the *P. E.* for a given value of r and n, but also

[1] The author has found the Gem adding machine (price $15, procurable through the Automatic Adding Machine Co., New York City, or through the C. H. Stoelting Co., Chicago) serviceable for work in which there is no necessity for printed records such as the Burroughs, Standard, Wales, and other high-priced machines afford.

determine the value of n, *i. e.*, the number of observations, needed to establish a given degree of correlation with any assigned degree of accuracy. In our illustrative case, since $n = 50$, and $r = .93$, we note that $P.\ E._r$ is less than 0.0181, that 200 observations would have reduced the error to less than .0091, etc. If our correlation had been lower, say 0.30, the error for 50 cases would have risen to 0.0868. Since the actual error is extremely small in relation to the obtained correlation, it follows that the latter has an enormous degree of reliability.

TABLE 5

Probable Error of r for Various Values of r and of n (Yule)

VALUES OF n	$r = 0$	$r = .1$	$r = .2$	$r = .3$	$r = .4$	$r = .5$	$r = .6$	$r = .7$	$r = .8$	$r = .9$	$r = 1$
25	.1349	.1335	.1295	.1228	.1133	.1012	.0863	.0688	.0486	.0256	.0000
50	.0954	.0944	.0916	.0868	.0801	.0715	.0610	.0486	.0343	.0181	.0000
75	.0779	.0771	.0748	.0709	.0654	.0584	.0498	.0397	.0280	.0148	.0000
100	.0674	.0668	.0648	.0614	.0567	.0506	.0432	.0344	.0243	.0128	.0000
200	.0478	.0473	.0459	.0435	.0402	.0359	.0306	.0244	.0172	.0091	.0000
300	.0389	.0386	.0374	.0354	.0327	.0292	.0249	.0199	.0140	.0074	.0000
400	.0337	.0334	.0324	.0307	.0283	.0253	.0216	.0172	.0121	.0064	.0000
500	.0302	.0299	.0290	.0274	.0253	.0226	.0193	.0154	.0109	.0057	.0000
600	.0275	.0273	.0264	.0251	.0231	.0207	.0176	.0140	.0099	.0052	.0000
700	.0255	.0252	.0245	.0232	.0214	.0191	.0163	.0130	.0092	.0048	.0000
800	.0238	.0236	.0229	.0217	.0200	.0179	.0153	.0122	.0086	.0045	.0000
900	.0225	.0223	.0216	.0205	.0189	.0169	.0144	.0115	.0081	.0043	.0000
1000	.0213	.0211	.0205	.0194	.0179	.0160	.0137	.0109	.0077	.0041	.0000

In general, a correlation, like any other determination, to have claim to scientific attention must be at least twice as large as its $P.\ E.$, and to be perfectly satisfactory, should be perhaps three to five times as large.

Since, in our illustration, r is some 51 times as large as its $P.\ E.$ its appearance by mere chance is practically zero and its reliability is practically absolute.

(b) The Pearson Method Adapted to Rank-Differences

The product-moments method gives full and exact weight to the d of each m from the M. If we disregard the magnitude of

these d's, however, and regard only the relative order or station of individual m's in each array, we may yet measure correlation by what is known as the method of rank-differences.

For this method, the formula is

$$r = 1 - \frac{\Sigma D^2}{c}, \tag{26}$$

in which D is the numerical difference between each corresponding pair of ranks[1] (not to be confused with d, the deviation from the mean), and in which c is the mean value of ΣD^2 by mere chance. Since

$$c = \frac{n(n^2 - 1)}{6}, \tag{27}$$

Formula 26 may be written:

$$r = 1 - \frac{6 \Sigma D^2}{n(n^2 - 1)}. \tag{28}$$

For illustration, note in Table 1, Column 11, the series of D's which are squared and summated in Column 12. Boy No. 30 ranks first (weakest) in the distribution for right-hand grip and first in the order for left-hand grip, hence his $D = 0$. Boy No. 52[2] is 3d in the first, and 9th in the second array, hence for him $D = 6$ and $D^2 = 36$. Since $n = 50$, by Formula 27,

$$c = \frac{50\ (2500 - 1)}{6} = 20{,}791;$$

hence, by Formula 26,

$$r = 1 - \frac{2098}{20{,}791} = 1 - .10 = .90,$$

[1] In this and the following rank method of Spearman, cases of 'ties' for a given rank are preferably divided in such a manner as to keep the total number of ranks equal in the two series. If, for instance, two S's rank 5th, they should both be assigned the rank 5.5 (to replace 5 and 6), or if three S's rank 5th, they should all three be assigned the rank 6 (to replace the 5th, 6th and 7th places in the series).

[2] To avoid possible confusion, it may be explained that two records were discarded, so that the boys' numbers run two over the fifty.

a result in close accordance with that of the product-moments method and obtained in a small fraction of the time.

(c) Spearman's Correlation 'Foot-Rule,' or R-Method.

Spearman's "'foot-rule' for measuring correlation" (17) is another and still simpler method of comparison by rank, the essential features of which are the use of D, the numerical difference of station, in place of D^2, and of only those of the D's that indicate a gain in rank (since the losses must equal the gains).

It is because this simplified method gives less weight to extreme measures that Spearman believes it to be actually more reliable for psychological purposes than the standard Pearson method. The advantages are summarized by Spearman (17, p. 104) as follows: "By using it, we obtain a precise quantitative value, which can be compared with that found by any other correlation under any circumstances or between any other things, either by the same or by the standard r-method; we free ourselves from various illusions, which are otherwise almost irresistible; we get a reliable estimate as to the danger of our result being merely an accidental coincidence; and we even learn how to plan out our experiments from the outset in a manner properly adapted to the object in view." In the opinion of Lehmann and Pedersen, however, Spearman's 'footrule' has only slight value, because it can measure and express only proportionality and not other uniform relationships of phenomena, while even when dealing with proportionality, the method may lead to totally false results, *e.g.*, with a complete inverse relation, the value secured, instead of -1 would be -0.5 with an odd number of cases, and $1-1.5 \times n^2 / (n^2-1)$ with an even number of cases (9, p. 18).

This method, it is important to note, yields an index, R, that is not identical with the Pearson r, though functionally related thereto, as is explained below.

The formula for Spearman's R is

$$R = 1 - \frac{\Sigma g}{c}, \tag{29}$$

in which g is the numerical gain in rank of an individual in the second, as compared with the first series, and in which c is the mean value of Σg by mere chance.

Since:

$$c = \frac{n^2 - 1}{6}, \tag{30}$$

we obtain by substitution:

$$R = 1 - \frac{6 \Sigma g}{n^2 - 1}. \tag{31}$$

For illustration, note in Table 1, Column 10, the series of gains (g), which yield $\Sigma g = 106$. As $n = 50$, by Formula 30, $c = (2500 - 1) \div 6 = 416.5$;
hence, by Formula 29,

$$R = 1 - \frac{106}{416.5} = 1 - .25 = .75.$$

To determine whether R has any claim to reliability, one may use the rather severe formula:[1]

$$P.E._R = \frac{0.43}{\sqrt{n}}. \tag{32}$$

To convert R-values into r-values we may use the formula:

$$r = \sin\left(\frac{\pi}{2} R\right); \tag{33}$$

or, for all cases in which R is less than .50, this may be simplified with little loss of accuracy into:

$$r = 1.5\, R. \tag{34}$$

For the quick and accurate conversion of R into r, Table 6, which is based on Formula 33, may be consulted.

In the correlation under examination, since $R = .75$, $r = .93$, or precisely the value obtained by the longer Pearson method.

(d) Correlation by Distribution of Selected Groups

The following method is sometimes useful as a device for preliminary survey, but when used, as it often has been, for a final

[1] According to Spearman, if we know that a correlation exists and wish merely to estimate the accuracy of R, a less rigorous formula may be used.

expression of correlation, it is inferior to the methods already described.

Distribute the data for both series in order as illustrated in Table 1, and divide them into four or five groups on the basis of equal numbers of cases or of equal amounts of deviation. By inspection it is often possible to determine at this juncture whether there is sufficient evidence of a correlation to justify further

TABLE 6

Conversion of R-Values into r-Values, in Accordance with Formula 33

R	r	R	r	R	r	R	r	R	r
.00	.00	.20	.31	.40	.59	.60	.81	.80	.95
.01	.01	.21	.32	.41	.60	.61	.82	.81	.96
.02	.03	.22	.34	.42	.61	.62	.83	.82	.96
.03	.05	.23	.35	.43	.62	.63	.84	.83	.96
.04	.06	.24	.37	.44	.64	.64	.84	.84	.97
.05	.07	.25	.38	.45	.65	.65	.85	.85	.97
.06	.08	.26	.40	.46	.66	.66	.86	.86	.98
.07	.11	.27	.41	.47	.67	.67	.87	.87	.98
.08	.13	.28	.43	.48	.69	.68	.88	.88	.98
.09	.14	.29	.44	.49	.70	.69	.88	.89	.99
.10	.16	.30	.45	.50	.71	.70	.89	.90	.99
.11	.17	.31	.47	.51	.72	.71	.90	.91	.99
.12	.19	.32	.48	.52	.73	.72	.90	.92	.99
.13	.20	.33	.50	.53	.74	.73	.91	.93	.99
.14	.22	.34	.51	.54	.75	.74	.92	.94	1.00
.15	.23	.35	.52	.55	.76	.75	.93	.95	1.00
.16	.25	.36	.54	.56	.77	.76	.93	.96	1.00
.17	.26	.37	.55	.57	.78	.77	.94	.97	1.00
.18	.28	.38	.56	.58	.79	.78	.94	.98	1.00
.19	.29	.39	.57	.59	.80	.79	.95	.99	1.00

calculation. For this inspection we may take the cases found in the first group of the first series and examine their distribution in the groups of the second series. Evidently, in the absence of correlation, these cases would be distributed by chance. Thus, in Table 1, let the two series be divided into 5 groups of 10 measurements each. The 10 measurements in the first group for right-hand grip would tend, by chance alone, to be distributed 2 in

each of the 5 groups in the second series, but as a matter of fact they are massed in or near the first group (8 in the 1st, 2 in the 2d), hence there is evidently a high degree of correlation.

The distribution in the second series of the remaining groups of the first series may be similarly tested, though an examination of the first group is commonly sufficient.

If the grouping is made in terms of deviation, the number of measurements found in the several groups will usually be unequal; it is then necessary to calculate the distribution of the various groups of the first series into those of the second. Suppose the two series of Table 1 are each divided into five 70-hectogram groups; the right-hand series will subdivide into groups containing 16, 15, 12, 4 and 3 measurements; the left-hand series into groups containing 12, 17, 12, 6 and 3 measurements, respectively. Take the 16 cases in the first group of the first series; by chance it is clear that 12/50, or 3.84 of them, would fall in the first group of the 2d series, 17/50, or 5.44 of them would fall in the second group, 3.84 in the third, 1.97 in the fourth, and .98 in the fifth. The actual distribution of the 16 cases into those five groups is 11, 5, 0, 0, 0, as compared with the chance distribution, 3.84, 5.44, 3.84, 1.97, .98.

If now we wish not only to explore the distribution of selected groups tentatively for the presence of correlation, but also to present the evidence of the correlation in compact form, we may, as is often done, prepare a table by comparative averages. To return again to the correlation of right and left-hand strength of grip, we may by this means secure the following tabular statement of the relations between right-hand and left-hand grip.[1]

TABLE 7

Correlation of Right and Left-Hand Grip by Group Averages (Whipple)

Group in 1st Series	1st 10	2d 10	3d 10	4th 10	5th 10
Average Right Hand Grip	194.9	229.9	271.5	317.9	402.2
Average Left Hand Grip	183.7	226.3	269.2	291.5	394.4

[1] For examples of this type of correlation, consult Bagley (1), Binet and Vaschide (2). Though this method has been frequently used, a little reflection will show that it is inferior to the several methods described before, because the advantage of weighing the relation of each individual measure is lost by lumping them into averages. and because, moreover, no coefficient of correlation is computed. If the groups were made more numerous and the data presented in the form of a graph showing the entire course of the relationship the absence of the coefficient might be less serious.

The data from such a table may be thrown into graphic form very simply: let ordinates represent the values of the one series, abscissæ the values of the other, so that the one series is plotted as a function of the other: if, then, the first series ranges upward for values from low to high and the second from left to right for values from low to high, a positive correlation will be indicated by a line running in a southwest-northeast direction, inverse correlation by a line running in a northwest-southeast direction, and zero correlation by a vertical or a horizontal line (depending on which series is plotted on the ordinates). In proportion as the correlation is complete the line assumes an oblique position.

(c) Correlation of Presence and Absence

Suppose we wish to correlate two traits whose presence or absence is ascertainable, but about whose degree of presence nothing can be said. Such a correlation can be computed if we first determine:

a = number of cases in which both traits are present,
b = number of cases in which the first trait is present and the second absent,
c = number of cases in which the second trait is present and the first absent,
d = number of cases in which both traits are absent.

On the basis of these values, we may use the simple formula of Yule:

$$r = \frac{ad - bc}{ad + bc}, \tag{35}$$

or, better, the following modification:

$$r = \sin \frac{\pi}{2} \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}. \tag{36}$$

If, in this formula, we replace the sine by the cosine of its complement, we secure

$$r = \cos \left[\frac{\pi}{2} - \frac{\pi}{2} \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}} \right],$$

which we can reduce to

$$r = \cos \frac{\sqrt{bc}}{\sqrt{ad} + \sqrt{bc}} \pi. \tag{37}$$

The probable error, provided ab is not very unequal to cd, may be taken as

$$P.E._r = \frac{1.1}{\sqrt{n}}. \tag{38}$$

As an example one may take the measure of the tendency of white cats to be afflicted with deafness (as cited by E. Davenport from Yule). From the data as presented in Table 8, we may calculate by Formula 35 that $r = .91$.

TABLE 8

Relation of Deafness and White Color in Cats (Yule)

CATS	WHITE	NOT WHITE
Deaf	6	4
Not Deaf	14	976

Now this same formula can often be used for preliminary exploration of series in which the degree of presence of a trait is known, and which may, therefore, be treated, if desired, by the more elaborate methods. For this purpose, assume that all measurements greater than M (or the median), *i. e.*, all *plus* cases, signify the presence of the trait, and all *minus* cases its absence. In Table 1 we find, using the median, 22 cases that are plus in both series, 3 that are plus in right and minus in left-hand grip, 3 that are minus in right and plus in left-hand grip, and 22 that are minus in both. Substituting these values in Formula 37, we have

$$r = \cos \frac{\sqrt{9}}{\sqrt{484} + \sqrt{9}} \pi, \text{ or } \cos 21.6° = .93,$$

which gives again the same value as by the standard formula

(f) The Method of Unlike Signs

Even this procedure may be simplified by substituting for $\sqrt{bc}$ the *percentage* of cases with unlike signs (U), and for $\sqrt{ad}$ the percentage of cases with like signs (L)[1] with the result,

$$r = \cos \frac{U}{L + U} \pi, \tag{39}$$

or, since $L + U$ must always equal 100, and since $\pi = 180°$, this formula may be condensed, if desired, to

$$r = \cos U \; 1.8°. \tag{40}$$

Finally, since U must lie between 50 and 0 for positive and between 50 and 100 for inverse correlations, a table may be prepared[2] from which the values of r may be read directly from any integer value of U.

By reference to the paragraph above it will be seen that in Table 1 we have 6 cases of unlike signs in the 50, hence $U = 12$ and $r = 0.93$, as by other methods.

TABLE 9

Corresponding Values of r and U for Formula 40 (Whipple)
If U is greater than 50, first subtract it from 100, then prefix the minus sign to the correlation indicated

U	r	U	r	U	r	U	r	U	r
0	1.000	10	.951	20	.809	30	.587	40	.309
1	.999	11	.941	21	.790	31	.562	41	.279
2	.998	12	.929	22	.770	32	.536	42	.248
3	.995	13	.917	23	.750	33	.509	43	.218
4	.992	14	.904	24	.728	34	.482	44	.187
5	.987	15	.891	25	.707	35	.454	45	.156
6	.982	16	.876	26	.684	36	.426	46	.125
7	.976	17	.860	27	.661	37	.397	47	.094
8	.968	18	.844	28	.637	38	.368	48	.062
9	.960	19	.827	29	.613	39	.338	49	.031

[1] That is, virtually substituting the arithmetical for the geometrical mean.

[2] Reproduced from an earlier article by the author (23) in which the applicability of the method is discussed more fully.

This method cannot be recommended for final determinations of important correlations because the probable error is too large, but it is a useful device for quick examination of a relation.

The method of unlike signs, in sharp contrast to the product-moments method, disregards entirely the amount of deviation. The author has made occasional use, for the first approximation, of an intermediate method that also disregards the actual deviations, but introduces artificial ones to be treated by the standard Pearson formula. Assume that all measurements lying between *M* (or the median) and the *A. D.* (or *S. D.*, or *P. E.*) have a deviation of 1 (plus or minus according to direction of the actual deviations), and that all other measurements have a deviation of 2 (plus or minus). Compute Σxy, σ_1 and σ_2 precisely as if these assumed deviations were the actual ones. This method, when applied to Table 1, yielded $r = .97$ when σ was calculated by Formula 7, and $r = .87$ when σ was calculated by Formula 5.

3. *The Correction of Obtained Correlations to their True Value*

(a) Correction of the Attenuation Produced by Chance Errors

The real correspondence between two traits or capacities is not, as has so often been erroneously supposed, necessarily revealed by the determination of a coefficient of correlation, even by the most approved methods and with a probable error that is satisfactorily small. All measurements, as we have noted, are subject to chance errors of observation. In the determination of averages such errors tend to counterbalance one another, so that if the measurements are sufficiently numerous, the obtained *M* differs from the true *M* by an inappreciable amount. In the case of correlations, however, these errors[1] are not eliminated by increasing the number of observations, and their presence has the effect of *decreasing* the size of the correlation, so that, in so far as these errors are concerned, the 'raw' or obtained correlation is too small, or, to use Spearman's term, the correlation is 'attenuated' by errors which constitute, from this point of view, constant or systematic errors.

[1] The phrase 'errors of observation' is to be understood in a wide sense, to include not only errors arising from technique, instrumentation, etc., but also *chance* shifts in the disposition of subjects, in their attitude toward the test, etc.

This illusory attenuation of the correlation by errors of observation seems, in fact, a principal cause of the contradictory nature of results that have hitherto been obtained; in experiments in which such errors have been very large, a correlation has not appeared, even when present, and has, in consequence, been erroneously denied. The determination of a small correlation therefore opens two possibilities; it may indicate actual absence of correspondence, or it may indicate merely the presence of large chance errors of observation. (Krüger and Spearman, p. 55.)

In order to correct the raw and discover the true r, it is imperative to secure at least two independent series of observations. The formula for correction of attenuation, or the 'expanding' formula, as it might be termed. is then applied as follows:

$$AB_t = \frac{M\ (A_1B_1,\ A_1B_2,\ A_2B_1,\ A_2B_2)}{M\ (A_1A_2,\ B_1B_2)} \tag{41}$$

in which

AB_t = the true correlation,
M = the mean,
A_1 = the 1st series of observations of the trait A,
A_2 = the 2d series of observations of the trait A,
B_1 = the 1st series of observations of the trait B,
B_2 = the 2d series of observations of the trait B,
A_1B_1 = the raw correlation of A_1 and B_1,
A_1A_2 = the raw correlation of A_1 and A_2, etc.

Thus the numerator is the M of the four possible r's between the measurements of A and the measurements of B, while the denominator is the M of the r of the two A series and the r of the two B series. Incidentally, these last mentioned correlations, taken singly, afford an obvious coefficient of reliability of the two series of measurements of A and B respectively.

The above formula holds for ordinary cases, but if one series of observations, say A, should be known to be much more exact and reliable than the other, then the geometrical should be substituted for the arithmetical M. In theory the denominator should always be the geometrical M, but the arithmetical M is virtually as accurate and for short series even more desirable. For the mathematical demonstration of this and the following formulas, the accuracy of which has been disputed by several writers, consult Spearman (15). The correction does not entirely eliminate the uncertainty that arises from the use of 'random samples' for investigation; that must be removed by the use of more extended series.

(b) Correction of the 'Constriction' or 'Dilation' Produced by Constant Errors

Attenuation is the result of the operation of chance errors,—chance in the sense that the deviation of any measurement takes place independently of the deviation of any other measurement. If, however, some influence is at work which affects all the measurements of one or of both series, such a constant factor or constant error will prove a source of disturbance that may either increase or decrease the obtained correlation. Such disturbances will result from the operation of any factor which is not strictly relevant to the correspondence under examination.

If an irrelevant factor affects both of the series, it is evident that the correlation will be unduly increased or 'dilated.' Suppose for example, that one wished to determine the correlation of pitch discrimination with the discrimination of lifted weights, and that the subjects of the experiments were of different ages. Then, since the two capacities in question both tend to improve with age, this common dependence on age will clearly tend to induce the appearance of a correlation, even if there really be none between the capacities themselves when compared under uniform conditions of age.[1]

If an irrelevant factor affects but one of the series, it is evident that the correlation will be unduly decreased or 'constricted,' *i. e.*, the irrelevant influence will tend to reduce any proportionality that really exists between the two series. To quote an example from Spearman, a correlation of 0.49 was discovered between pitch discrimination and school standing, but it was likewise discovered that more than half the children had 'taken lessons,' and thus had the opportunity for special training in the observation of pitches.

These constant irrelevant factors may not always be excluded,

[1] This undiscovered or neglected influence of age has been a very common source of error in many studies of correlation. Obviously, this particular irrelevancy may be eliminated practically by proper selection of subjects for the investigation, or it may be eliminated by manipulation of the results in various ways besides that here described: see, for example, Bagley.

but their force can frequently be measured and allowed for by the following formula:[1]

$$AB_t = \frac{AB_a - AC \cdot BC}{\sqrt{(1 - AC^2)\,(1 - BC^2)}}, \tag{42}$$

in which

AB_t = the true correlation between A and B,
AB_a = the apparent correlation between A and B.
AC = the direct correlation between A and any irrelevant factor, C,
BC = the direct correlation between B and C.

If, as is most often the case, the irrelevant factor affects but one series, this influence of 'constriction' may be excluded by the simpler formula:

$$AB_t = \frac{AB_a}{\sqrt{(1 - AC^2)}} \tag{43}$$

Thus, in the example mentioned, the correlation between pitch discrimination and its disturbing factor, musical training, was found by computation, to be 0.61; hence, by Formula 43,

$$r = \frac{0.49}{\sqrt{1 - 0.61^2}} = 0.62.$$

From the above considerations, it follows that the experimenter must define with some exactness the traits that are to be examined for a possible correlation, and that he must not seek to establish the correlation until, by means of suitable preliminary exploration, he has discovered all the irrelevant factors that might disturb the correspondence. The mere mechanical computation of an index of correlation does not, then, demonstrate the existence of a real correlation, or at least, does not accurately and certainly define its nature. Hence, while, as we have seen, we may very hopefully look to correlational work for revelation of the functional disposition of mind, this is no royal road to the attainment of the end, but can itself be entered upon in each instance only after a preparatory survey and critical inspection of the problem in hand has afforded sufficient acquaintance with the traits and capacities that are therein concerned. One must be a psychologist as well as a statistician.[2]

---

[1] AB, AC and BC must first be 'expanded' by Formula 41.

[2] The study of the functional correlation of five well-known tests by Krüger and Spearman affords an admirable illustration of the value of such a combination of sound psychology and sound statistics.

4. The Discovery of Common Factors in Intercorrelated Capacities

If three or more psychological traits show intercorrelations one with another, the question may be raised as to whether the intercorrelations are not due to the presence of some common factor to which all the capacities are functionally related, whether, in other words, these correlations may not arise from a single underlying cause. If such a common or 'central' factor be assumed to be present, we may test the validity of the assumption by mathematical procedure, leaving the exact nature of the factor out of consideration for the time being. For example, if for any given capacity, A, we have obtained two independent measurements, A_1 and A_2, and if for two other capacities, we have obtained the measurements B and C respectively, then the correlation (AF) between the capacity A and the hypothetical common or central factor, F, may be determined by the formula:

$$\mathrm{AF} = \frac{M\ (\mathrm{AB},\ \mathrm{BC})}{M\ (\mathrm{A_1A_2},\ \mathrm{BC})}. \tag{44}$$

In illustration, Krüger and Spearman found the following values,—correlation of pitch discrimination with the Ebbinghaus completion test, 0.65, with adding, 0.66, correlation of two measurements of pitch discrimination, 0.87, correlation of the Ebbinghaus test with adding, 0.66; hence the correlation of pitch discrimination with the hypothetical central factor is the M of 0.65 and 0.66 divided by the M of 0.87 and 0.71, or 0.83.[1]

REFERENCES [2]

(1) W. C. Bagley, On the correlation of mental and motor ability in school children, in A. J. P., 12: 1901, 193–205.

(2) A. Binet and N. Vaschide, Expériences de force musculaire et de fond chez les jeunes garçons in A. P., 4: 1897 (1898), 15–63.

---

[1] This central factor is tentatively ascribed by these authors to some psychophysiological condition, possibly general neural plasticity. In another article Spearman speaks of "General Discrimination" or "General Intelligence," or the "Intellectual Function." Here, again, is revealed the instructive wealth of possibilities in the application of adequate methodological treatment to psychic life.

[2] Consult the *List of Abbreviations* for the exact titles of periodicals in these and subsequent references.

(3) C. B. Davenport, Statistical methods, with special reference to biological variation, N. Y., 1899. Pp. 148. (2d revised edition, 1904.)

(4) E. Davenport, Principles of breeding; a treatise on thremmatology, Boston, 1907. Pp. 713. (With an appendix on statistical methods by H. Rietz.)

(5) E. Dürr, review of Spearman (15, 16), in Z. P., 41: 1906, 450. Also, Erwiderung, in Z. P., 42: 1906, 470–2.

(6) W. P. Elderton, Frequency curves and correlation, London 1906.

(7) F. Galton, Natural inheritance, London, 1889. Pp. 259.

(8) F. Krueger and C. Spearman, Die Korrelation zwischen verschiedenen geistigen Leistungsfähigkeiten, in Z. P., 44: 1907, 50–114.

(9) A. Lehmann and R. Pedersen, Das Wetter und unsere Arbeit, in A. G. P., 10: 1907, 1–104, especially 15–26.

(10) M. Merriman, A text-book on the method of least squares, N. Y., 1884. Pp. 194.

(11) M. Meyer, The grading of students, in Science, 28: 1908, 243–250.

(12) H. Rietz: see E. Davenport.

(13) E. C. Sanford, A course in experimental psychology, Boston, 1895 and 1898. Pp. 449.

(14) E. W. Scripture, The new psychology, London, 1897. Pp. 500.

(15) C. Spearman, The proof and measurement of association between two things, in A. J. P., 15: 1904, 72-101.

(16) C. Spearman, General intelligence objectively determined and measured, in A. J. P., 15: 1904, 201-293.

(17) C. Spearman, 'Footrule' for measuring correlation, in B. J. P., 2: 1906, 89–109.

(18) C. Spearman, Entgegnung, in Z. P., 42: 1906, 467–470.

(19) C. Spearman, Demonstration of formulæ for true measurements of correlation, in A. J. P., 18: 1907, 161–9.

(20) E. L. Thorndike, Educational psychology, N. Y., 1903. Pp. 173.

(21) E. L. Thorndike, An introduction to the theory of mental and social measurements, N. Y., 1904. Pp. 210.

(22) E. B. Titchener, Experimental psychology, vol. ii, Quantitative experiments, Parts I and II, N. Y., 1905.

(23) G. M. Whipple, A quick method for determining the index of correlation, in A. J. P., 18: 1907, 322–5.

(24) C. Wissler, The correlation of mental and physical tests, in P. R. M. S., 3: No. 6, 1901. Pp. 62.

THE TESTS

CHAPTER IV

Anthropometric Tests

The tests embraced in this chapter have been developed primarily as anthropometric tests. They do not include tests of physical capacity or function (Chapter V), but simply measurements of bodily size or dimension.

The number of such measurements that have been made and recorded runs well into the hundreds, and an extensive literature has appeared. The science of anthropometry has developed partly in connection with anthropology and sociology, partly in connection with the study of physical development, including bodily growth, hygiene, gymnastic and athletic training. In recent years, moreover, a not inconsiderable contribution has been made by psychologists, physicians, educators, and other investigators who have been interested in the correlation between bodily and mental traits.

It is this last-mentioned phase of anthropometry that concerns us, and hence only a few important measurements that have assumed special importance in conjunction with other physical and with mental tests are here considered.

The references which follow will enable the reader to study the development of anthropometry and the application of anthropometric tests at large. Bertillon and Galton should be consulted by those who are interested in the use of anthropometric measurements in the identification of criminals: Key and Hertel have given special consideration to the relation of growth to disease and to hygienic conditions. Anthropometric charts or record-books have been published by E. Hitchcock. D. A. Sargent, J. W. Seaver, W. W. Hastings, Anna Wood, L. H. Gulick, and others.

REFERENCES

(1) A. Bertillon, Signaletic instructions; including the theory and practise of anthropometrical identification, Eng. Trans., Chicago and New York, 1896. Pp. 260 + 81 plates, charts, etc.

(2) H. G. Beyer, The growth of United States naval cadets, in Proc. U. S. N. Inst., 21: No. 2, 1895.

(3) F. Boas, The correlation of anatomical or physiological measurements, in Amer. Anthropologist, 4: 1894, 313.

(4) F. Boas, The growth of children, in Science, 19: 1892, 250.

(5) F. Boas, On Dr. W. T. Porter's investigation of the growth of the school children of St. Louis, in Science, 1: 1895, 225.

(6) F. Boas, Growth of first-born children, in Science, 1: 1895, 402.

(7) F. Boas, Growth of Toronto children, in Rept. Brit. Ass. Adv. Science for 1897, 1898, 443.

(8) F. Boas, Growth of American children, in U. S., 1896–7, ii, 1555.

(9) H. P. Bowditch, The growth of children, in 10th An. Rept. State Brd. Health, Mass., 1879, 33.

(10) H. P. Bowditch, The growth of children, studied by Galton's method of percentile grades, in 22d An. Rept. State Brd. Health, Mass., 1891, 479.

(11) F. Burk, Growth of children in height and weight, in A. J. P., 9: 1897–8, 253.

(12) A. Engelsperger and O. Ziegler, Beiträge zur Kenntnis der physischen und psychischen Natur des sechsjährigen in die Schule eintretenden Kindes. Anthropometrisches Teil, in E. P., 1: 1905, 173–235.

(13) Lucy Ernst and E. Meumann, Das Schulkind in seiner körperlichen und geistigen Entwicklung. Part 1, by L. Ernst, Leipzig, 1906. Pp. 143.

(14) W. Farr, *et al.*, Table showing the relative statures of boys at the age of 11 to 12 years under different social and physical conditions of life, in Rept. Brit. Ass. Adv. Science, 1880, i, 127.

(15) F. Galton, Proposal to apply for anthropological statistics from schools, in J. Anthrop. Inst., 3: 1873–4, 308.

(16) F. Galton, On the height and weight of boys aged 14 in town and country schools, in Nature, 5: 1875–6, 174.

(17) F. Galton, Range in height of males at each age and in several classes, in Rept. Brit. Ass. Adv. Science, 51: 1882, 250.

(18) F. Galton, Head growth in students at the University of Cambridge, in Nature, 38: 1888–9, 14.

(19) F. Galton, Useful anthropometry, in Proc. Amer. Ass. Adv. Phys. Educ., 6: 1891, 51.

(20) F. Galton, Anthropometrical instruments, in Anthrop. J., 16: 2.

(21) F. Galton, Anthropometric percentiles, in Nature, 31: 1884–5, 223.

(22) F. Galton, Finger-prints, London, 1892. Pp. 216.

(23) A. Gihon, Physical measurements, in Wood's Ref. Hand-bk. Med. Sciences, 5: 1887, 667.

(24) N. A. Gratsianoff, Data for the study of physical development in childhood (in Russian), St. Petersburg, 1889.

(25) J. M. Greenwood, Heights and weights of children, in Public Health Ass. Rept., 1891, Concord, N. H., 17: 1892, 199.

(26) E. Hartwell, Anthropometry in the United States, in Amer. Statis. Ass., 3: 554.

(27) E. Hartwell, Rept. of the director of physical training, Boston Normal Schools, in School Doc., No. 8, Boston, 1894.

(28) E. Hartwell, Rept. of the director of physical training, in School Doc., No. 4, Boston, 1895.

(29) E. Hartwell, Bowditch's law of growth and what it teaches, reprint from 10th An. Proc. Amer. Ass. Adv. Phys. Educ., Concord, N. H., 1896.

(30) W. Hastings, A manual for physical measurements for use in normal schools, public and preparatory schools, etc., Springfield, Mass., 1902. Pp. 112.

(31) A. Hertel, Neuere Untersuchungen über den allgemeinen Gesundheits-Zustand der Schüler u. Schülerinen, in Schulgesundheitspflege, Nos. 6 and 7, 1888, 167–183, 201–215. (Contains a review of the work of Key.)

(32) E. Hitchcock, Jr., Physical measurements, fallacies, and errors, in Proc. Amer. Ass. Adv. Phys. Educ., 1887, 35.

(33) E. Hitchcock, Jr., A synoptic exhibit of 15,000 physical examinations, Ithaca, 1890.

(34) E. Hitchcock, Comparative study of measurements of male and female students at Amherst, Mt. Holyoke, and Wellesley Colleges, in Proc. Amer. Ass. Adv. Phys. Educ., 6: 1891, 37.

(35) E. Hitchcock, The results of anthropometry as derived from the measurements of the students of Amherst College, Amherst, 1892.

(36) E. Hitchcock and H. Seelye, An anthropometric manual, giving the average and mean physical measurements and tests of male college students, and modes of securing them, Amherst, 1889.

(37) B. Holmes, A study of child growth, being a review of the work of Dr. Wm. T. Porter, of St. Louis, in N. Y., Med. J; 60: 1894, 417.

(38) W. Jackson, Jr., Graphic methods in anthropometry, in Phys. Educ., 2: 1893, 89.

(39) A. Key, Schulhygienische Untersuchungen, trans. and edited, somewhat condensed, from the Swedish, by L. Burgerstein, Hamburg and Leipzig, 1899. (Report of the Swedish Commission of 1882 to investigate hygienic conditions of the schools.)

(40) A. MacDonald, Experimental study of children, including anthropometrical and psychophysical measurements of Washington school children, reprint of chs. 21 and 25 of U. S., 1897–8, Washington, 1899.

(41) P. Malling-Hansen, Perioden im Gewicht der Kinder und in der Sonnenwärme, Copenhagen, 1886.

(42) S. Moon, Measurements of the boys of the McDonogh School, McDonogh, Md., 1893.

(43) K. Pearson, Growth of St. Louis children, in Nature, 51: 1894, 145–6.

(44) G. Peckham, The growth of children, in Rept. Wis. Brd. Health, 1881, 28.

(45) G. Peckham, Various observations on growth, *ibid.*, 1882, 185.

(46) W. T. Porter, On the application to individual school children of the mean values derived from anthropological measurements by the generalizing method, in Pub. Amer. Statis. Ass., n. s. 3: 1892–3, 576.

(47) W. T. Porter, The growth of St. Louis children, in Trans. Acad. Science, St. Louis, 6: 1894, 263.

(48) W. T. Porter, Physical basis of precocity and dullness, *ibid.*, 6: 1893–4, 160.

(49) W. T. Porter, The relation between growth of children and their deviation from the physical type of their sex and age, *ibid.*, 233.

(50) W. T. Porter, Use of anthropometrical measurements in schools, in Educ. Rev., 11: 1896, 126–133.

(51) C. Roberts, A manual of anthropometry.

(52) Sack, Physical development of the children in the middle schools of Moscow, 1892.

(53) D. Sargent, Report on anthropometric measurements in Proc. Amer. Ass. Adv. Phys. Educ., 2: 1886, 6.

(54) D. Sargent, Anthropometric apparatus, with directions for measuring and testing the principal physical characteristics of the human body, Cambridge, Mass., 1887.

(55) D. Sargent, The physical proportions of the typical man, in Scribner's Magazine, 2: 1887, 3.

(55) D. Sargent, The physical development of women, *ibid.*, 5: 1889, 541.

(56) D. Sargent, Strength-tests and strong men at Harvard, in J. Boston Soc. Med. Sci., No. 13, 1896–7.

(57) G. Schultz, Some new anthropometrical data, in Yale Med. J., 2: 1895–6, 149.

(58) J. Seaver, Anthropometry and physical examination, New Haven, 1890.

(59) G. Sergi, An anthropological cabinet for pedagogic purposes, in Education, 7: 1886, 42–9.

(60) C. Stratz, Der Körper des Kindes, für Eltern, Erzieher, Aerzte, u. Künstler, Stuttgart, 1904.

(61) F. Swain, Anthropometric measurements, in Proc. Amer. Ass. Adv. Phys. Educ., 3: 1887, 43.

(62) G. Tarbell, On the height, weight, and relative growth of normal and feeble-minded children, in Proc. 6th An. Session Ass. Med. Officers Amer. Inst. Idiotic and Feeble Minded Persons, Phil., 1883, 188.

(63) R. Thoma, Untersuchungen über die Grösse u. das Gewicht der anatomischen Bestandtheile des menschlichen Körpers, 1882.

(64) E. B. Titchener, Anthropometry and experimental psychology, in Philos. Review, 2: 1893, 187–192.

(65) F. Tuckermann, Anthropometric data based upon nearly 3,000 measurements taken from students, Amherst, 1888.

(66) G. West, Worcester school children; the growth of the body, head, and face, in Science, 21: 1893, 2–4.

(67) G. West, Observations on the relation of physical development to intellectual ability, made on the school children of Toronto, Canada, in Science, n. s. 4: 1896, 156.

(68) C. Wissler, Growth of boys; correlations for the annual increments, in Amer. Anthropol., n. s. 5: 1903, 81.

(69) M. Wood, Anthropometric tables, arranged after the method of percentile grades, of the measurements of 1500 Wellesley students.

(70) M. Wood, Anthropometric tables, compiled from the measurements of 1100 Wellesley College students, 1890.

(71) M. Wood, Statistical tables concerning the class of 1891 of Wellesley College, numbering 104 women.

(72) M. Wood, Statistical tables, showing certain measurements of forty freshmen of Wellesley College (before and after gymnasium training), 1892.

TEST I.

Height, standing and sitting.—The general purpose of this test is, of course, to furnish a measurement of height as an index of physical size or growth for the sake of comparison with mental traits or with other physical traits. It is included in practically every series of tests that include any physical measurements.

APPARATUS.—Stadiometer (Fig. 3). Small calipers (Fig. 4) or millimeter rule.

METHOD.—(1) For standing height, the examiner, *E*, should, when feasible, have the subject, *S*, remove his shoes, and stand on the stadiometer with the heels together and with heels, buttocks, the spine between the shoulders, and the head, all in contact with the measuring rod. The chin must not be unduly raised or depressed. *E* then brings down the sliding arm of the instrument until it rests squarely, but without excessive pressure, upon *S*'s head.

(2) For sitting height, let *S* sit erect upon the stand of the stadiometer with spine and head in contact with the measuring rod.

RESULTS.—(1) The best norms of stature are doubtless those calculated by Boas (3)[1] from studies by various investigators

[1] The figures in parentheses following names refer to the reference-numbers at the end of the test in which they occur.

of school children (45,151 boys and 43,298 girls) in Boston, St. Louis, Milwaukee, Toronto, and Oakland, Cal.[1] For the sake of comparison with these norms and with the norms for strength

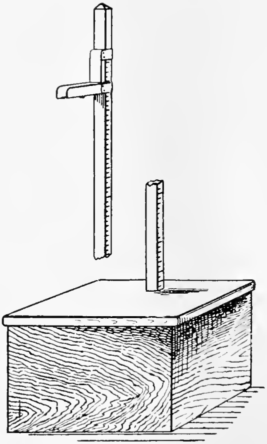


FIG. 3. STADIOMETER, OR HEIGHT STAND.

Graduated in tenths of inches on one side and millimeters on the other.

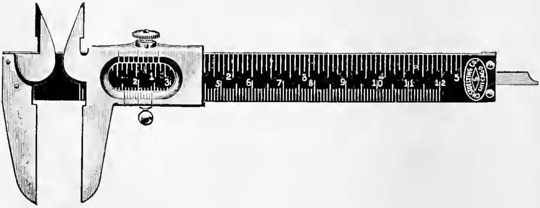


FIG. 4. VERNIER CALIPER, FOR EXTERNAL, INTERNAL, AND DEPTH MEASURING.

Fitted with both English and metric scales and verniers for each, reading to $\frac{1}{128}$ of an inch and $\frac{1}{10}$ of a millimeter.

---

[1] The same averages converted into inches may be found in Burk, while these and other studies are summarized by MacDonald. Consult Boas for table showing the distribution of stature at each age according to the frequency method. Valuable tables and charts showing the distribution of height and of other anthropometric measurements by percentile grades will be found in Smedley (17).

of grip, vital capacity, etc., to be quoted later, there are given herewith the norms of standing and sitting height derived from the measurement of 2788 boys and 3471 girls by Director Smedley of the Department of Child-Study and Pedagogic Investigation, Chicago (16).

TABLE 10

*Norms of Stature of American Children, in cm. (Boas)**

Age	5.5	6.5	7.5	8.5	9.5	10.5	11.5
Boys	105.90	111.58	116.83	122.04	126.91	131.78	136.20
Girls	104.88	110.08	116.08	121.21	126.14	131.27	**136.62**

Age	12.5	13.5	14.5	15.5	16.5	17.5	18.5
Boys	140.74	146.00	152.39	159.72	164.90	168.91	171.07
Girls	**142.52**	**148.69**	**153.50**	156.50	158.03	159.14	——

*The figures in black-faced type in Tables 10–14 indicate periods in which the averages for girls exceed those for boys of the same ages. The rapid growth of puberty and early adolescence is initiated and terminated earlier in girls than in boys.

TABLE 11

Norms of Standing and Sitting Height, in cm. (Smedley)

AGE	STANDING HEIGHT		SITTING HEIGHT		AGE	STANDING HEIGHT		SITTING HEIGHT	
	BOYS	GIRLS	BOYS	GIRLS		BOYS	GIRLS	BOYS	GIRLS
6.0	110.69	109.66	62.40	61.72	12.5	141.89	**144.32**	74.70	**76.29**
6.5	113.25	112.51	63.54	62.90	13.0	145.54	**147.68**	76.24	**77.91**
7.0	115.82	115.37	64.67	64.07	13.5	149.09	**151.04**	77.79	**79.54**
7.5	118.39	118.22	65.78	65.25	14.0	151.92	**153.64**	79.21	**80.99**
8.0	120.93	120.49	66.75	66.34	14.5	154.74	**156.24**	80.64	**82.43**
8.5	123.48	122.75	67.72	67.43	15.0	158.07	156.83	82.18	**83.21**
9.0	126.14	125.24	68.79	68.32	15.5	161.41	157.42	83.68	83.99
9.5	128.80	127.74	69.85	69.21	16.0	164.03	158.30	85.43	84.54
10.0	130.91	130.07	70.56	70.05	16.5	166.65	159.18	87.17	85.09
10.5	133.03	132.41	71.26	70.89	17.0	167.85	159.26	88.16	85.20
11.0	135.11	**135.35**	72.10	**72.23**	17.5	169.04	159.34	89.14	85.30
11.5	137.19	**138.30**	72.93	**73.58**	18.0	171.23	159.42	90.30	85.51
12.0	139.54	**141.31**	73.80	**74.93**	18.5	173.41	159.50	91.46	85.72

From these and other statistics, the following important results may be gathered:

(2) There is a period of slower growth in height in boys at 11 years of age, and a similar, though less marked, retardation in girls at nine years of age.

(3) During the period of approximately 11 to 14 years girls are taller than boys of the same age, because the prepubertal acceleration of growth occurs earlier in girls.

(4) Sitting-height follows the same general laws as standing-height.

(5) Boys continue their growth in height later than do girls, *i. e.*, maturity in height is not reached so early.

(6) Children of purely American descent are taller than those of foreign-born parentage (Bowditch, Peckham).

(7) Children of the non-laboring classes are as a group taller than children of the laboring classes (Bowditch, Roberts).

(8) According to Bowditch, large children make their most rapid growth at an earlier age than small ones, but according to Boas (1, 2) this induction is untenable.

(9) The height of American-born children is modified by density of population. Urban life decreases stature from five years of age (Peckham, 10, 11).

(10) According to Kline, boys in the public schools are taller than boys in truant schools, save at the age of ten. Similarly, Smedley (17) found the boys in the Chicago School for incorrigibles and truants shorter than normal boys from the tenth year up.

(11) Gratsianoff and Sack in Russia, and Porter (39–40), MacDonald and Smedley (17) in America, have concluded that bright children are taller than dull children. West (18), however, found exactly the opposite to be true, while Gilbert (6, 7) found no constant relation between height and mental ability. Porter and Smedley determined mental ability by the relation of grade and age, Gilbert and MacDonald by the teacher's estimate.

(12) Children with abnormalities are inferior in height to children in general (MacDonald).

NOTES.—The upright measuring rod should be braced in such a manner that it will not be bent out of place by the pressure of *S*'s back. Many *S*'s will be inclined to assume an unnatural

position in this examination, especially to stretch themselves: the apparent height may be increased by as much as 20 to 30 mm. in this way.

If it is not practicable to remove the shoes, height may be taken with them on, and the height of the heel may subsequently be determined by the use of the small calipers or millimeter rule, and then subtracted from the gross height; the resulting error will be very small.

Height, as is well known, decreases slightly during the day, owing to the packing of the intervertebral cartilages and the loss of muscular tone: this loss in height during the day amounts, in the case of young men, to from 10 to 18 mm. It is therefore desirable, for accurate work, to take height measurements at approximately the same period of the day. It might be possible to work out empirically a corrective formula.

Porter's correlation between height and mental ability raises an important question which reappears whenever we discuss the correlation between any physical trait, *e. g.*, weight, strength, vital capacity, etc., and mental ability. The trend of evidence is to the effect that all such correlations, where found, are largely explicable as phenomena of growth, *i. e.*, as correlations with relative maturity (Cf. Boas, 2; Wissler, 19). This makes intelligible the fact that, in general, the positiveness of all such correlations lessens with age, and that many of them, indeed, become difficult or impossible of demonstration in adults. Thus, to take the correlation in question, a positive correlation is not, of course, to be interpreted as meaning that, taken individually, all tall boys are, *ipso facto*, bright boys, but that, taken collectively, those boys whose physical condition is good, whose growth is unimpaired by ill-health, faulty nutrition, etc., and who realize to the full the possibility of physical development inherent in them (whether they will ultimately be short or tall) will be found to exhibit the best mental condition and the most rapid mental development.

REFERENCES

(1) F. Boas, The growth of children, in Science, 19: 1892, 250.

(2) F. Boas, On Dr. W. T. Porter's investigation of the growth of the school children of St. Louis, in Science, n. s. 1: 1895, 225.

(3) F. Boas, Growth of American children, in U. S., 1896–7, ii. 1555.

(4) H. P. Bowditch, The growth of children, studied by Galton's method of percentile grades, in 22d An. Rept. State Brd. Health Mass., 1891, 479.

(5) F. Burk, Growth of children in height and weight, in A. J. P., 9: 1897–8, 253.

(6) J. A. Gilbert, Researches on the mental and physical development of school children, in Yale S., 2: 1894, 40–100.

(7) J. A. Gilbert, Researches upon school children and college students, in Iowa S., 1: 1897, 1–39.

(8) L. W. Kline, Truancy as related to the migratory instinct, in Pd. S. 5: 1898, 381–420.

(9) A. MacDonald, Experimental study of school children, etc., in U. S., 1899 (chs. 21 and 25.).

(10) G. Peckham, The growth of children, in Rept. Wis. Brd. Health, 1881, 28.

(11) G. Peckham, Various observation on growth, *ibid.*, 1882, 185.

(12) W. T. Porter, The growth of St. Louis children, in Trans. Acad. Science, St. Louis, 6: 1894, 263.

(13) W. T. Porter, Physical basis of precocity and dullness, *ibid.*, 6: 1893–4, 160.

(14) C. Roberts, A manual of anthropometry.

(15) Sack, Physical development of the children in the middle schools of Moscow, 1892.

(16) F. W. Smedley, Rept. of the dept. of child-study and pedagogic investigation, reprint from 46th An. Rept. Brd. Educ., Chicago, 1899–1900. Also in U. S., 1902, i., 1095–1115.

(17) F. W. Smedley, do., No. 3, 1900–1901, also in U. S., 1902, i., 1115–1138.

(18) G. West, Observations on the relation of physical development to intellectual ability, made on the school children of Toronto, Canada, in Science, n. s., 4: 1896, 156.

(19) C. Wissler, The correlation of mental and physical tests, in P. R. M. S. 3: No. 6, 1901. Pp. 62.

TEST 2

Weight.—The general purpose of determining weight is similar to that of determining height, viz: to furnish an index of physical size or growth as a basis for correlation with other tests or observations.

Apparatus.—Accurate scales, preferably of the type especially devised for anthropometric work, which allow readings to be rapidly and accurately taken in the metric system, with units of 50 g. or twentieths of a kilogram (Fig. 5). If avoirdupois

scales are used, they should be divided into tenths of pounds rather than into ounces.

METHOD AND TREATMENT OF RESULTS.—For accurate measurements, weight should be taken without clothes. Where this is impracticable, the weight of the clothes may be deducted by subsequent measurement. For some comparative purposes, however, the weight of the clothes may be neglected and the figures obtained

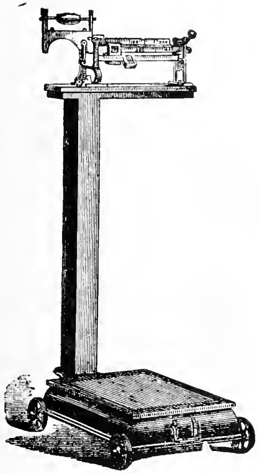


FIG. 5. ANTHROPOMETRIC SCALES.

The platform side of the beams is graduated metric to 100 kilos, by 50-gram divisions, and the other side avoirdupois to 200 pounds, by tenths of a pound.

from the gross weight may be taken for computation, or these figures, better yet, may be corrected by arithmetical computation based upon the weights of the clothes of a limited number of *S*'s.

We may form a tolerably accurate notion of the 'clothing error' by reference to investigations upon this point. Thus, according to W. S. Christopher (3), who ascertained the weight of the ordi-

nary schoolroom clothing of 121 Chicago children, chiefly in the month of May, "the average weight of the clothing of all the pupils was 5.5 per cent of the gross weight" (boys, 5.8 per cent; girls, 5.2 per cent.) These figures varied little with age: obese children wore clothing lighter in proportion to their weight than that worn by others, while "the most variable element in the clothing was found to be the shoes, especially the shoes worn by the boys." Only a few children wear clothing that weighs more than 7 per cent, or less than 4 per cent of their gross weight.

RESULTS.—(1) From the data of about 68,000 children in the cities of Boston, St. Louis, and Milwaukee, Burk (2) derives the norms reproduced in Table 12: the Chicago norms are reproduced in Table 13.

TABLE 12

Norms of Weight, in kg. (Burk)

Approx. Age	6.5	7.5	8.5	9.5	10.5	11.5	12.5	13.5	14.5	15.5	16.5	17.5
Boys	20.50	22.45	24.72	27.03	29.66	32.07	34.88	38.46	43.18	48.72	54.88	—
Girls	19.69	21.64	23.81	26.03	28.53	31.52	**35.70**	**40.23**	**44.59**	48.40	50.94	52.34

TABLE 13

Norms of Weight, in kg., with Clothing (Smedley)

Age	6.0	7.0	8.0	9.0	10.0	11.0	12.0	13.0	14.0	15.0	16.0	17.0	18.0
Boys.....	19.738	21.613	23.817	26.336	28.707	31.223	34.151	38.084	42.696	47.993	53.238	57.384	61.283
Girls	18.870	20.974	23.010	25.257	27.795	30.662	**34.373**	**38.974**	**44.219**	**48.161**	50.652	52.386	52.923

(2) As in the case of height, girls exhibit the prepubertal increase in weight some two years earlier than boys, and are for the years 12 to 15 heavier than boys of the same age.

(3) Growth in weight, as in height, is subject to some lessening of rate at 9 years for girls and at 11 for boys.

(4) Boys continue to increase in weight after girls have practically attained their maximal normal weight. Girls grow most rapidly from 10 to 15 years, boys from 12 to 17 years.

(5) Mean variations in weight are largest during the period of

fastest growth, which shows that not all individuals participate equally or evenly in the rapid growth of adolescence.

(6) First-born children exceed later-born children in weight, at least during the period from 6 to 15 years, though the reverse is true of the weight at birth. The difference is slight, but very regular (Boas, 1).

(7) Children of the non-laboring classes are as a group heavier than children of the laboring classes (Bowditch).

(8) Children of American-born parents are heavier than those of foreign-born parents.

(9) The correlation between weight and mental ability or precocity is found to be positive by some investigators, negative by others, and indifferent by still others. Thus, Porter (8) asserts very positively that "precocious children are heavier and dull children lighter than the mean child of the same age," and draws a further practical conclusion that "no child whose weight is below the average for its age should be permitted to enter a school grade beyond the average of its age, except after such a physical examination as shall make it probable that the child's strength be equal to the strain." Porter's conclusions are confirmed by Smedley (9) at Chicago. On the basis of the teacher's estimate of mental ability, Gilbert (4, 5), however, finds no constant relation between weight and such ability, save that from 10 to 14 years the dull children are much heavier than the bright, while West (10), who used a similar basis, finds a negative correlation throughout.

(10) Both Kline (6) and Smedley (9) find the mean weight of boys in truant schools to be less than that of boys in the public schools, save at the age of 10.

(11) Porter concludes that the acceleration in weight preceding puberty takes place at the same age in dull, mediocre, and precocious children, but investigations in New York City seem to oppose this conclusion and indicate rather that puberty and pubertal growth is distinctly earlier in precocious children, *i. e.*, that mental and physical precocity go hand in hand.

(12) Children with abnormalities are below the average in weight (MacDonald, 7).

Notes.—It is not important to have scales which render possible a very fine measurement, such as fractions of an ounce, because the

normal weight of any individual varies from day to day and from hour to hour during the day: the daily variation is, in the case of young men, as high as 0.3 kg. The author, in a long series of observations conducted at the same hour daily, found gains and losses of more than 1 kg. in 24 hours. It may not be amiss in this connection to point out the absurdity of attaching any significance to small gains or losses that are observed in weighings conducted at occasional and irregular intervals. Severe exercise may reduce the weight by a large amount; *e.g.*, two hours of football practise may take off 2 or 3 kg. from a man who is not yet in training. It is well, however, for comparative purposes, to take weight measurements at approximately the same period of the day.

REFERENCES

(1) F. Boas, Growth of first-born children, in Science, n.s. 1: 1895, 402-4.

(2) F. Burk, Growth of children in height and weight, in A.J.P., 9: 1897-8, 253.

(3) W. S. Christopher, Report on child-study, reprint from An. Rept. Brd. Educ. Chicago, 1898-1899.

(4) J. A. Gilbert, Researches on the mental and physical development of school children, in Yale S., 2: 1894, 40-100.

(5) J. A. Gilbert, Researches upon school children and college students, in Iowa S., 1: 1897, 1-39.

(6) L. W. Kline, Truancy as related to the migratory instinct, in Pd.S., 5: 1898, 381-420.

(7) A. MacDonald, Experimental study of school children, etc., reprint of chs. 21 and 25 of U. S., 1899.

(8) W. T. Porter, Physical basis of precocity and dullness, in Trans. Acad. Science St. Louis, 6: 1894, 263.

(9) F. Smedley, Rept. dept. child-study and pedagogic investigation, No. 3, in Rept. Brd. Educ. Chicago, 1900-1901, also in U. S., 1902, i., 1115-1138.

(10) G. West, Observations on the relation of physical development to intellectual ability, made on the school children of Toronto, Canada, in Science, n.s. 4: 1896, 156.

TEST 3

Diameter of the skull.—This measurement has been commonly conducted for the purpose of investigating the correlation between size of the head and general intelligence. It forms also one of the chief measurements undertaken in the Bertillon system

for the identification of criminals. The following directions are adapted from Bertillon's account (1).

A. MEASURING THE LENGTH OF THE HEAD

INSTRUMENT.—Head calipers (Fig. 6).

METHOD.—(1) Seat *S* with his right side toward a window, and stand facing his left side. Hold the left tip of the calipers firmly in place at the glabella (space between the eye-brows) with the

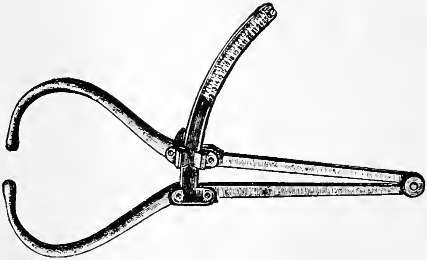


FIG. 6. HEAD CALIPERS.

tip of the instrument between the thumb and forefinger, and with these resting on the adjacent parts of the forehead to prevent the compass-tip from deviating.

(2) Hold the calipers in an approximately horizontal plane so that the scale is fully lighted by the window, with the right tip projecting about one cm. beyond the finger-tips of the right hand. Keep the eyes fixed upon the scale; then bring the right tip down over the back and middle of the head until it has passed the most projecting point; then move the tip upward again, making sure that it is well within the hair and in constant contact with the scalp; continue these exploring movements so as to pass the maximal point two or three times, keeping the eyes constantly fixed upon the scale to detect this point.

(3) Remove the calipers and set them by tightening the set-screw at the supposed length; take care to set them accurately within 0.5 mm.

(4) Replace the calipers thus set and tightened, and again execute the exploring movements described in (2). If the setting is correct, the instrument will just *touch* the skin of the head at the maximal point, but will pass over it without undue friction and without necessitating pressure upon its arms: one millimeter too short will produce definite resistance at this point; one millimeter too long, a definite lack of friction. Practise will enable E to distinguish the 'feel' of the correctly set instrument, and errors should not exceed 1 mm.

B. MEASURING THE WIDTH OF THE HEAD

INSTRUMENT.—Head calipers as above.

METHOD.—Position of S, preliminary exploring movements, setting of the calipers, and subsequent verification follow the same general procedure as in the determination of the length of head. The following additional instructions are to be noted:

(1) E stands behind S, and is careful to preserve an erect, symmetrical position, in order to ensure equal freedom with both elbows and a symmetrical position of the calipers.

(2) Hold the calipers a short distance from each end; apply the tips first at the upper point of attachment of each ear; then raise them *vertically* and watch the scale to determine the point of greatest width, making several testing movements both upward and downward.

(3) The true maximal diameter in most cases is not yet found, but lies in the same horizontal plane as the preliminary maximal point just determined, and about 3 cm. behind it. Hence, next move the calipers slowly back and forth two or three times in a *horizontal* plane and determine the true maximal point.

(4) Set the instrument, as in the previous measurement, and verify the setting. In this verification, the caliper-points should describe a series of zig-zag movements, in order certainly to traverse the areas of maximal width (usually less than the size of a dime), which might not be traversed if the movements were circular or too coarsely executed.

TREATMENT OF RESULTS.—From the measurements of the length and width of head, the *cephalic index* may be computed readily by

multiplying the width by 100 and dividing by the length. This index is considered one of the most important of those used in anthropometry. By it, the type of head may be determined as follows: if the index is less than 75, *S* is long-headed (dolichocephalic); if 75–80.9, *S* is 'medium' (mesocephalic); if 81–86.9, *S* is broad-headed (brachycephalic); if 87 or over, *S* is excessively broad-headed (hyperbrachycephalic).

RESULTS.—(1) Typical head measurements are those made by Boas, West, Chamberlain and others upon Worcester school children, and reported by West (7): these are reproduced in Table 14.

TABLE 14

Diameters of the Skull and the Cephalic Index (West)

AGE	AVERAGE LENGTH		AVERAGE WIDTH		CEPHALIC INDEX	
	Boys	Girls	Boys	Girls	Boys	Girls
	mm.	*mm.*	*mm.*	*mm.*	%	%
5	176	174	140	138	79.56	79.40
6	177	172	142	139	78.94	79.60
7	179	175	142	140	79.42	80.02
8	180	174	143	141	78.71	80.41
9	181	176	144	140	79.63	79.71
10	182	177	145	142	80.30	79.46
11	183	180	144	142	78.80	78.90
12	183	180	145	143	79.40	79.40
13	184	181	147	145	79.50	79.60
14	187	183	147	144	78.60	79.00
15	188	184	148	146	78.59	78.99
16	191	184	149	144	77.81	78.48
17	189	185	150	146	78.34	78.50
18	192	186	151	147	78.88	79.36
19	192	183	150	145	78.33	79.68
20	195	182	152	147	77.88	79.41
21	192	186	153	145	79.29	78.36

(2) In boys, length of head continues to increase until the age of 21; in girls, maximal length of head is practically attained at 18. The growth, both of length and width of head, is very irregular, *i.e.*, periods of growth alternate with periods of cessation of growth.

(3) Boys' heads are longer and wider than those of girls throughout the whole period of growth, and consequently throughout life

(4) Width of head is greater in precocious than in dull children (Porter, 6, 7). One method of arraying data to show this principle is illustrated in Table 15, in which all the girls aged 12 and all the boys aged 10 are distributed according to their school grades. It is then seen that those children of a given age in an advanced school grade have, on the average, broader heads than those in a lower grade.

TABLE 15

Breadth of Head by School Grade (Porter)

SCHOOL GRADE	BOYS AGED 10		GIRLS AGED 12	
	Cases	Average	Cases	Average
		mm.		*mm.*
I............	92	145.86	—	—
II............	408	146.73	68	143.68
III............	397	146.48	193	144.77
IV	170	147.21	343	144.94
V............	—	—	217	145.50
VI............	—	—	89	147.64

(5) Binet, somewhat similarly, finds that the head of the unintelligent is smaller than that of the intelligent child in all dimensions save in vertical diameter and distance from the base of the nose to the end of the chin, though the differences are but slight and somewhat uncertain. If, however, exceptionally bright children (*enfants d'élite*) are compared with exceptionally dull children (*enfants arrièrès*), differences averaging 3–4 mm. or more appear, particularly in transverse dimensions, *i.e.*, in breadth. Exceptionally bright children distinctly surpass average children, but the latter do not differ so much from dull children. In brief, then, exceptionally bright children are characterized by unusually wide heads.

(6) According to MacDonald, "dolichocephaly increases in children as ability decreases. A high percentage of dolichocephaly is, to a certain extent, a concomitant of mental dullness." "Unruly boys have a large percentage of long-headedness."

(7) The measurements, by Engelsperger and Ziegler, of 238 boys and 238 girls of the entering classes (average age 6 years, 4.5

months) in the schools of Münich, furnish results that deviate somewhat from those just cited for American children, as is seen clearly by a comparison of Tables 14 and 16.

It is of interest to note that no cases of dolichocephaly were found, but that these children were decidedly brachycephalic.

(8) Miss Lee found no correlation between the estimated skull capacity and the intellectual capacity of 60 men and 30 women.

NOTES.—Heads of unusual shape or size, irregular or deformed, should receive especial care in measurement, and a descriptive note should be appended to the record.

Attempts to record the shape and size of the skull by means of the registering 'conformateur' used by hatters have usually been relinquished, because the hair interferes too much with exact de-

TABLE 16

Skull Dimensions and Proportions of Entering Classes at Münich (Engelsperger and Ziegler)

	LENGTH			BREADTH			INDEX		
	Mean	Max.	Min.	Mean	Max.	Min.	Meso.	Brachy.	Hyperb.
							%	%	%
Boys	170.35	185	148	146.34	160	133	6.3	54.6	39.1
Girls	165.83	186	151	142.97	159	130	6.7	45.8	47.5

termination. This instrument might, however, be of service in preserving a rough 'picture' of heads of unusual size or proportions.

Measurement of the head was formerly regarded as a very obvious means for the estimation of the size and proportions of the brain, and hence of intellectual ability, but it is easy to demonstrate that these relations obtain only in the gross. That is, brain size and form are only roughly indicated by the exterior dimensions of the head, while intelligence is conditioned primarily by the elaborateness of the finer nerve structure and not (save in pathological cases of hypertrophy or developmental arrest) by the gross size or form of the brain. It is scarcely necessary to call attention to the absurdity, *a fortiori*, of the claims of phrenology.

REFERENCES

(1) A. Bertillon, Signaletic instructions; including the theory and practise of anthropometrical identification, Eng. tr., Chicago and New York, 1896. Pp. 260.

(2) A. Binet, Recherches sur la technique de la mensuration de la tête vivante, and four other articles on cephalometry, in A. P., 7: 1900 (1901) 314-429.

(3) A. Engelsperger and O. Ziegler, Beiträge zur Kenntniss der physischen und psychischen Natur des sechsjährigen in die Schule eintretenden Kindes. Anthropometrisches Teil in E. P., 1: 1905, 173-235.

(4) Alice Lee, Study of the correlation of the human skull, in Science, n. s. 12: 1900, 946-9.

(5) A. MacDonald, Experimental study of children, etc., reprint of chs. 21 and 25 of U. S., 1899.

(6) W. T. Porter, The growth of St. Louis children, in Trans. Acad. Science St. Louis, 6: 1894, 263.

(7) W. T. Porter, Physical basis of precocity and dullness, *ibid.*, 160.

(8) G. West, Worcester school children; the growth of the body, head, and face, in Science, 21: 1893, 2-4.

TEST 4

Girth of the skull.—This measurement is less in favor with investigators than those just described, because of the variable factor of the hair, just mentioned.

INSTRUMENT.—Anthropometric measuring tape (Fig. 7).

METHOD.—*E* stands at the right of *S*, who is seated. *E* holds

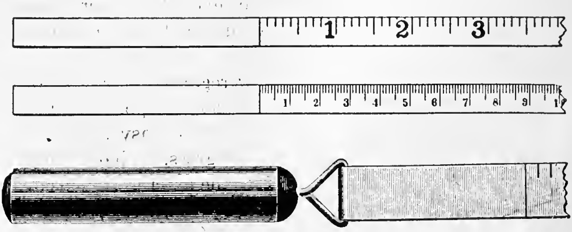


FIG. 7. ANTHROPOMETRIC TAPE.

Graduated to tenths of inches on one side and millimeters on the other, and fitted with a spring handle to eliminate the personal equation of different examiners.

the tape with the thumb and forefinger of each hand at a length approximately that of the distance to be measured. He then lifts the tape over *S*'s head, keeping it horizontal, and applies it about the head at such a height as to pass around the largest part—over the frontal prominences and over the occipital prominences. The tension of the tape is regulated by observation of the spring-indicator.

RESULTS.—(1) Measurements of the circumference of the head of 7953 boys and 8520 girls in Washington, D.C., by MacDonald (3)[1] form the basis for the results embodied in Table 17.

(2) The head circumference of boys is larger than that of girls save in the case of colored children. (Colored girls have a larger circumference of head at all ages than white girls.) In American

TABLE 17

Circumference of the Head, in Inches (MacDonald)

AGE	BOYS	GIRLS	AGE	BOYS	GIRLS
6	20.48	20.20	13	21.01	20.95
7	20.45	19.94	14	21.21	21.18
8	20.51	20.14	15	21.45	21.28
9	20.61	20.29	16	21.67	21.38
10	20.73	20.43	17	21.87	21.55
11	20.82	20.54	18	21.91	21.60
12	20.94	20.78			

children, the measurements of girls most nearly approach those of boys at 13 and 14, or at the period when the girls excel in height and weight.

(3) Children of the non-laboring classes have a larger circumference of head than children of the laboring classes.

(4) Children with abnormalities are inferior in head circumference to normal children.

(5) In an examination of 60 juvenile delinquents, Dawson (2) found the average circumference of head less than that of normal children of the same age: in 64 per cent of the cases studied, the

[1]Consult MacDonald (pp. 1016ff.) for an extended discussion of the relation of circumference of head to sex, nativity, race, sociological condition and mental ability.

circumference was from 1.7 to 5.2 cm. less than the mean for normal children.

(6) According to MacDonald, as circumference of head increases, mental ability (as reported by the teacher) increases, provided that one and the same race be under consideration. Möbius (4), similarly, asserts that, at least in the case of normal (*gesunde*) adults, mental capacity tends to exhibit correlation with skull capacity. Bayerthal (1) measured the skull circumference of 234 boys and 153 girls (ages 7.5 to 8.5 years) and related these measures with school standing by classifying both sexes into five groups, as "very good" (I), "good" (II), "good on the whole" (III), "satisfactory" (IV), and "more or less unsatisfactory" (V). The results tend to confirm the existence of a positive correlation between skull circumference and general ability. Thus, the average skull-circumferences were, for boys, 51.46, 50.93, 50.33, 49.60, and 49.60 cm., and for girls, 50.00, 49.83, 49.44, 49.16, and 48.84 cm., for the groups I to V, respectively.

The same investigator found that in one class of 48 girls, who had been classed by their teacher into the three groups, good, average, and poor, the average skull circumference for these groups was 49.6, 49.16, and 48.75 cm., respectively.

REFERENCES

(1) Bayerthal, Kopfumfang und Intelligenz im Kindesalter, in E. P., 2: 1906, 247-251.

(2) G. E. Dawson, A study in youthful degeneracy, in Pd. S., 4: 1896, 221-258.

(3) A. MacDonald, Experimental study of school children, etc., reprint of chs. 21 and 25 of U. S., 1899.

(4) Möbius, Geschlecht und Kopfgrösse, Halle, 1903.

CHAPTER V

Tests of Physical and Motor Capacity

The title 'physical and motor capacity' is here used as a conveni-ent and practical phrase to cover a number of tests which have often been classified under diverse rubrics, such as strength tests, motor tests, physical tests, tests of physiological condition, etc. All of the tests here described differ from the anthropometric tests of Chapter IV in that they measure not mere size or dimension, but functional, especially muscular, capacity. They differ from the tests of Chapter VI, many of which might equally well be said to measure physiological condition or capacity, *e.g.*, the test of visual acuity, in that they are primarily tests of motor, rather than of sensory capacity.

The first test described, that of vital capacity (often loosely termed lung capacity) is, perhaps, not so obviously a test of muscular efficiency as are the four strength tests that follow. It is, however, clearly a test of physical capacity dependent upon movement. The tests of quickness, accuracy, and steadiness of movement are frequently placed in a class by themselves under the rubric 'motor tests,' but they are easily subsumed under the caption here employed.

Reaction-time would by many be considered a test of quickness of movement; but it is so largely dependent upon complex psychological conditions, particularly upon the instructions, the direction of attention, and the type of stimulus employed, that it belongs rather to the experimental examination of action than to the measurement of physical capacity as such.[1]

These tests of physical and motor capacity have become prominent chiefly because of their employment in the study of the corre-

[1] See an article by the writer, "Reaction-Times as a Test of Mental Ability," in A. J. P., 15: 1904, 489.

lation of physical and mental ability. For this purpose they are commonly used in conjunction with the anthropometric tests already described and with various tests of general intelligence or mental ability to be described later.

These tests have also an obvious and direct application in the study of various problems of hygiene, physical culture, etc.

TEST 5

Vital capacity.—Vital capacity, also termed breathing capacity and differential capacity, is the maximal volume of air that can be

FIG. 8. WET SPIROMETER.

Graduated in cubic inches and cubic decimeters.

expired after taking a maximal inspiration. It is not identical with lung capacity, because a certain amount of air, termed the residual air, always remains in the lungs.

Vital capacity is considered an important index of general physi-

cal condition and capacity, and has, accordingly, found a place in nearly all measurements of school children in which the physical status has been examined. It is affected by sex, age, stature, posture, occupation, amount of daily physical activity, and by disease, and may be markedly increased (*e.g.*, 300 cc., in three months) by various forms of physical exercise which demand active respiration.

The ratio of vital capacity to weight is termed the *vital index* and is held to be of extreme significance, because it expresses the balance between bodily size and the rate and completeness with which oxidization of the blood is, or may be, effected. A high vital index is undoubtedly a preventive of auto-intoxication, gives increased resistance to disease, and is the root of endurance under effort. Thus athletic training consists primarily in the reduction of weight and the increase of breathing capacity.

TABLE 18

Norms of Vital Capacity, in Cubic Centimeters (Smedley)

AGE	BOYS	GIRLS	AGE	BOYS	GIRLS
6	1023	950	13	2108	1827
7	1168	1061	14	2395	2014
8	1316	1165	15	2697	2168
9	1469	1286	16	3120	2266
10	1603	1409	17	3483	2319
11	1732	1526	18	3655	2343
12	1883	1664			

APPARATUS.—Spirometer, preferably of the wet type (Fig. 8) fitted with detachable wooden mouth-piece. Extra mouth-pieces.

METHOD.—See that *S*'s clothing is perfectly loose about his neck and chest. Instruct him to stand upright, to take as full an inspiration as possible, and then to blow, not too rapidly, into the spirometer. Also caution him to take care that no air escapes about the mouth-piece.

Two or three trials may be allowed, and the best record set down.

After *S*'s record is made, discard the mouth-piece and insert a new one into the rubber tube.

RESULTS.—(1) The norms of vital capacity embodied in Table 18 are those established by Smedley (6) at Chicago.

(2) The relation between weight and vital capacity, *i.e.*, *the vital index*, presented in Table 19, is that found by Kotelmann (4), also given by MacDonald (5). The ratio expresses the relation in terms of kg. of weight and cc. of vital capacity. It will be seen that the weight of the body normally increases with age somewhat faster than the vital capacity. If height be similarly treated, it will be found, on the contrary, that vital capacity increases with age faster than it increases.

(3) All investigators agree that boys have a larger vital capacity than girls at all ages, and that men, similarly, have a larger capacity than women. Even if we compare men and women of the same height, the former surpass the latter by about the ratio 10 : 7.5.

TABLE 19

Value of the Vital Index, when Weight is Taken as Unity (Kotelmann)

AGE	INDEX	AGE	INDEX	AGE	INDEX	AGE	INDEX
9	69.32	12	67.51	15	63.18	18	64.28
10	69.37	13	66.75	16	65.94	19	66.22
11	69.18	14	64.07	17	65.77	20	65.01

(4) The average capacity for adults is computed by Vierordt to be for men 3400 cc. (a figure noticeably lower than the Chicago norms for boys in later adolescence) and for women 2400 cc.

(5) The norm is conditioned by height. For each centimeter of increase or decrease of stature above or below the mean, there is a corresponding rise or fall of the vital capacity, amounting in men to 60 cc., in women to 40 cc. This correlation with height varies somewhat at different ages. Thus, according to Wintrich, the average vital capacity for each centimeter of height is from eight to ten years, 10 cc., from 16 to 18 years, 20.65 cc., and at 50 years, 21 cc.

(6) In boys, growth in vital capacity is slow and steady during the years 6 to 12, but very marked during the next four years, whereas in girls the most rapid increase is during the years 11 to 14. In both sexes, these periods of rapid growth coincide with the periods of rapid growth in height and weight.

(7) Beyer (1), from his study of naval cadets, concludes that the maximal vital capacity is reached at 19, but other authorities place the maximum at 35, with an annual decrease of about 32 cc. thereafter, up to the age of 65.

(8) The most marked individual differences appear at the time when the period of most rapid growth terminates (Smedley, 7).

(9) Vital capacity is proportionately reduced in men who live a sedentary life. It is also reduced by any circumstance which interferes with the free expansion of the thorax, such as tight clothing, tuberculosis of the lungs, visceral tumors, etc.

(10) The correlation of vital capacity and mental ability is indifferent or negative according to Gilbert (2, 3), who found no constant relation, save that from 10 to 15 years duller children have the larger capacity. On the other hand, Smedley (6, 7) found a positive correlation between school standing and vital capacity, whether he took the distribution through the grades of all pupils of a given age, or computed the average school-grade of those who stood at various precentiles of vital capacity, or compared those at and above grade with those below grade at each age. Moreover, the same investigator found that pupils in the John Worthy School (incorrigibles, truants, etc.) were, from the age of ten up, inferior in vital capacity to children in the other schools, and that the inferiority became more noticeable with age.

NOTES.—The dry spirometer is less expensive than the wet in first cost, and is more portable, but it has the disadvantage of getting out of repair easily. Its readings are apt to run slightly higher than those of the wet spirometer.

The mouth-piece of the ordinary spirometer forms an excellent medium for the dissemination of bacteria. For this reason the detachable mouth-pieces are imperative if hygienic conditions are to be assured.

There is a certain knack in making a maximal spirometer record; some children may exhibit it; others not. In particular, to get a good record, the expiration must be neither too fast nor too slow, and an extra effort must be made just at the end of both inspiration and expiration to utilize the available lung-capacity to the utmost.

REFERENCES

(1) H. G. Beyer, The growth of United States naval cadets, in Proc. U. S. N. Inst., 21: No. 2, 1895.

(2) J. A. Gilbert, Researches on the mental and physical development of school children, in Yale S., 2: 1894, 40-100.

(3) J. A. Gilbert, Researches upon school children and college students, in Iowa S., 1: 1897, 1-39.

(4) L. Kotelmann, Die Körperverhältnisse der gelehrten Schüler des Johanneums in Hamburg, in Zeits. d. Königl. Preus. statist. Bureaus, 1879.

(5) A. MacDonald, Experimental study of school children, etc., reprint of chs, 21 and 25 of U. S., 1899.

(6) F. Smedley, Rept. dept. child-study and pedagogic investigation, reprint from 46th An. Rept. Brd. Educ. Chicago, 1899-1900. Also in U. S., 1902, i., 1095-1115.

(7) F. Smedley, do., No. 3, 1900-1901. Also in U. S., 1902, i., 1115-1138.

TEST 6

Strength of grip.—This test has been used to secure an index of general bodily strength, to secure an index of righthandedness[1] (in conjunction with Tests 10 to 12), and for comparative purposes generally. It may be modified to secure an index of endurance or fatigue (Test 9), or combined with other forms of strength measurement (Tests 7 and 8).

Apparatus.—Improved form of Smedley's dynamometer (Fig. 9). Millimeter rule.

Method.—With the millimeter rule, measure the distance from where *S*'s thumb joins his hand to the end of his fingers. Adjust the dynamometer by whirling the inner 'stirrup' until the scale on the outer stirrup indicates one-half this distance. This should bring the second phalanx to bear against the inner stirrup, and will ordinarily prove to be the optimal adjustment; if not, it may be modified to suit *S*'s inclinations. Then set the instrument by means of the clutch, so that the inner stirrup cannot twist while in use, and record the adjustment by reference to the scale upon the stirrup.

[1] The terminology of right and left-handedness is at present somewhat confused (Cf. E. Jones, in P. B., 6: April, 1909). The terms 'index of unidexterity' and 'index of dextrality' have been used by some writers as equivalent to 'index of righthandedness.' 'Dextrality' is here used to indicate the superiority of one hand (whether right or left) over the other.

Illustrate the use of the instrument to S: especially make clear that the lower pointer will register the grip, so that he does not have to continue his effort while the scale is read.

Allow three trials with each hand, right and left alternately, but introduce a brief pause, say 10 sec., between each trial to avoid excessive fatigue. Have S exert his maximal grip, and in each trial encourage him to do his best. Record the amount registered at each trial; but, for ordinary purposes, use in subsequent computation only the highest record for each hand.

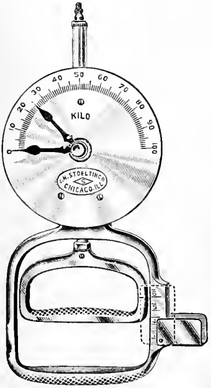


FIG. 9. DYNAMOMETER AND DYNAMOGRAPH, AFTER SMEDLEY, IMPROVED.

RESULTS.—(1) Tests of 2788 boys and 3471 girls in Chicago (9) with the Smedley dynamometer yielded the norms of Table 20.

(2) Boys are uniformly stronger than girls, and men stronger than women.

(3) The divergence between the *sexes* becomes marked at puberty, with the appearance of other sex traits.

(4) *Individual variation* in strength is more evident in early adolescence than at any other time.

(5) In his study of Washington school-children, MacDonald (6) found no correlation between *strength of hand and mental ability*,

but rather a dependence of strength upon sociological condition, *i.e.*, children of the poorer classes work outside of school hours and thus develop their strength. Since from these and other causes, the percentage of dull children in such a group is liable to be large, this accounts for the indications in his results that dull children tend to surpass bright or average children in strength. Yet Schuyten (7) concludes that the children of well-to-do parents are stronger than the children of poor parents.

TABLE 20

Norms of Strength of Grip, in kg. (Smedley)

AGE	BOYS		GIRLS	
	Rt. Hand	Lt. Hand	Rt. Hand	Lt. Hand
6	9.21	8.48	8.36	7.74
7	10.74	10.11	9.88	9.24
8	12.41	11.67	11.16	10.48
9	14.34	13.47	12.77	11.97
10	16.52	15.59	14.65	13.72
11	18.85	17.72	16.54	15.52
12	21.24	19.71	18.92	17.78
13	24.44	22.51	21.84	20.39
14	28.42	26.22	24.79	22.92
15	33.39	30.88	27.00	24.92
16	39.37	36.39	28.70	26.56
17	44.74	40.96	29.56	27.43
18	49.28	45.01	29.75	27.66

On the other hand, Miss Carman (3), from measurements of the grip of 1507 boys and girls aged 10 to 19 years, found that bright children exceeded dull children by an average of 3 kg. with the right, and 1 kg. with the left hand.

In Chicago (9, 10) the existence of a positive correspondence between strength of grip and class-standing was shown by three different methods, viz: by the distribution of 12-year old pupils by grades, by comparing the grip of those at and above grade with the grip of those below grade at each age, and by computing the average number of school grades that had been made by the various

percentile groups (in strength), after sex and age had been eliminated.

Schuyten, who estimated intelligence by school grade in relation to age, also found that those who are most intelligent are strongest.

(6) Dawson (5) found that juvenile *delinquents* have a mean strength of grip slightly less than normal children and that 56 per cent of them are inferior to the normal by from 1.32 to 11.82 kg. Similarly, boys in the school for incorrigibles and truants at Chicago are, at every age from 9 to 17 and with either hand, less strong than normal boys and this discrepancy increases very decidedly with age, *e.g.*, from 96.8 per cent of the norm at the age of 9 to 63.2 per cent of the norm at the age of 17.

(7) The *index of righthandedness, i.e.*, the percentage of strength of the left hand compared with the right, will be found to range, for any ordinary group of school children, between 91 and 96 per cent.

(8) *Dextrality, i.e.*, superiority of one hand over the other, is evident when the child enters school, but becomes increasingly evident as maturity approaches, and especially at puberty, so that a heightened difference in the strength of the hands may be regarded as one of the characteristic indications of pubertal change.

(9) There is a positive correlation between *dextrality and intellectual ability* (Smedley, 9), *i.e.*, dull pupils are more nearly ambidextrous than average, and average than bright, pupils, while the John Worthy schoolboys are still more nearly ambidextrous than the dull pupils of the regular schools.

(10) The *degree of dextrality* is greatest in the strongest children and least in the feeble, so that the latter may be said, as it were, to have two left hands (Binet and Vaschide, 1).

(11) If the test is taken under *stimulating conditions*, such as competition, personal encouragement, public announcement of records, etc., Binet and Vaschide found that the average grip was increased about 3 kg., or so much that the left hand surpassed the previous record of the right hand made without such incitement. Similarly, Schuyten (8) found that ennui, or loss of interest in successive tests, is sufficient to obscure the fatigue-effect of a school session.

(12) The exertion of maximal strength is commonly accom-

panied by characteristic poses, attitudes, facial contortions, grimaces, etc., which are, in general, evidences of the escape of uncontrolled energy through various motor paths. There appears to be an inverse relation between the strength and efficiency of the subject and the number and extent of these waste movements; these are correspondingly more evident when the muscles tire and *S* is unable to accomplish what he is attempting. In particular, a sort of foolish laugh is characteristic of this muscular inefficiency.[1]

NOTES.—The chief objections which have been made to the employment of the dynamometer are (1) that it is painful, particularly if a series of grips is taken, (2) that some *S*'s suffer from sweating of the hands, especially when excited, and that this causes the instrument to slip in their grasp, (3) that a wrong manner of holding the instrument may reduce the record, *e.g.*, by as much as 10 kg., (4) that, owing to the large number of muscles concerned, a lack of proper coördination in their contraction may lower the record.

The painfulness of the dynamometer can be largely eliminated by proper construction; the Smedley instrument is much better than the Collin elliptical form so commonly used heretofore. Moreover, if an extended investigation is to be undertaken, inurement to the pressure is rapidly developed (Bolton and Miller).

For the ascertainment of strength of grip, excessive perspiration can be avoided by simply drying the hands with a towel whenever necessary.

The proper holding of the instrument is also largely dependent on proper construction, and in this respect, again, the Smedley instrument, with its adjustable grip, is a distinct improvement over other forms.

The last objection is not to be seriously considered, first, because hand-grip is one of the most common forms of coördinated movement and is well organized early in childhood, and second, because experience shows that most *S*'s can make their maximal record in three attempts at least.[2]

[1] For a description, with photographic reproductions of these motor automatisms of effort, consult Binet and Vaschide.

[2] If, for any reason, *E* considers these sources of error not eliminated, it may be necessary to select a number of *S*'s and coach them in the use of the dynamometer until they can either avoid the errors, or report to *E* when they occur. J. Clavière (4) asserts that to employ only those *S*'s who are thus trained in the use of the instrument is an indispensable condition for successful dynamometry.

In any extended investigation, *E* should take steps to test the calibration of the dynamometer occasionally. For this purpose, the instrument is held securely in a vise or other support and a series of weights are hung upon the stirrup while the scale-readings are compared with the actual weighting.

For many purposes it is desirable to combine strength of grip with strength of back and strength of legs by adding the data secured in these three tests.

The advantages and disadvantages of using a series of grips in place of a single one are discussed in Test 9.

REFERENCES

(1) A. Binet and N. Vaschide, Expériences de force musculaire et de fond chez les jeunes garçons, in A. P., 4: 1897 (1898), 15–63. See also pp. 173–199, 236–252, 295–302.

(2) T. Bolton and Eleanora Miller, On the validity of the ergograph as a measurer of work capacity, in Nebraska Univ. Studies, 1904. Pp. 79 + 128.

(3) Ada Carman, Pain and strength measurements of 1507 school children in Saginaw, Michigan, in A. J. P., 10: 1899, 392–8.

(4) J. Clavière, Le travail intellectuel dans ses rapports avec la force musculaire mesurée au dynamomètre, in A. P., 7: 1900 (1901), 206-230.

(5) G. E. Dawson, A Study in youthful degeneracy, in Pd. S., 4: 1896, 221–258.

(6) A. MacDonald, Experimental study of school children, reprint of chs. 21 and 25, U. S., 1899.

(7) M.-C. Schuyten, Les variations de la force musculaire et le développement intellectuel des élèves; summarized by A. Binet, in A. P., 9: 1902 (1903), 448–9, from Pædologisch Jaarboek, Ghent, 1902.

(8) M.-C. Schuyten, Comment doit on mesurer la fatigue des écoliers, in Ar. P., 4: 1904, 113–128.

(9) F. Smedley, Rept. dept. child-study and pedagogic investigation, reprint from 46th An. Rept., Brd. Educ. Chicago, 1899–1900. Also in U. S., 1902, i., 1095–1115.

(10) F. Smedley, do., No. 3, 1900–1901. Also in U. S., 1902, i., 1115–1138.

TEST 7

Strength of back.—This test, together with the following, has been extensively used in securing an index of the general bodily strength of college students, but has not been applied in most examinations of school children. A fairer index of strength may, however, be gained by its use in combination with strength of grip.

INSTRUMENT.—Back and leg dynamometer (Fig. 10).

METHOD.—*S* stands upon the footrest of the instrument, which *E* should then adjust by lengthening or shortening the chain, so that *S*'s body is inclined forward at an angle of about 60 degrees (Fig. 11). *S* should then take a full breath and give a hard lift, mostly with the back and without bending the knees. Two or three trials may be recorded, and the best record used subsequently in computation.

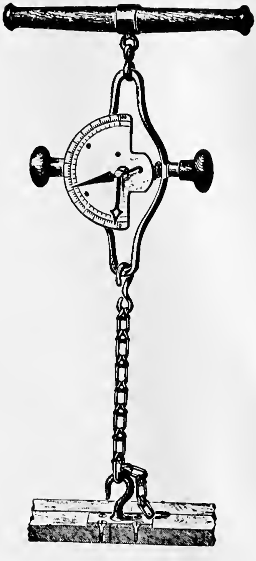


FIG. 10. BACK AND LEG DYNAMOMETER. CAPACITY, 700 KG.

RESULTS.—(1) On the use of this and the succeeding test, with quantitative results as obtained in college gymnasiums, etc., consult Hastings (3), Sargent (4, 5), Seaver (6) and other authorities already cited under anthropometry in general.

(2) Binet and Vaschide (1) found the lift (*force renale*) of 37 boys aged from 12 to 14 years to average 77 kg., with a maximum of

121 kg., and minimum of 56 kg. With 40 young men averaging 18 years of age, the same investigators (2) obtained for the average 146.64, for the maximum 187, and for the minimum 101.6 kg. Hastings (p. 71) publishes measurements of 5000 young men (17-30 years) whose strength of back averages 150.9 kg., *P. E.* 22.1, with a minimal record of 74.5 and a maximal record of 227.3.

(3) Back lift is roughly about 3.2 times the strength of the right hand.

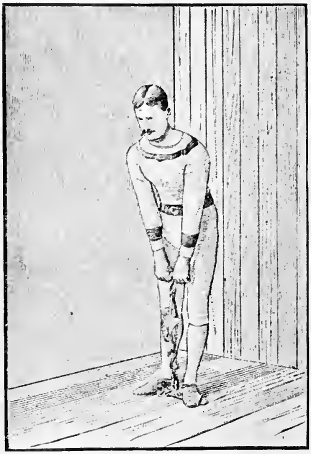


FIG. 11. BACK AND LEG DYNAMOMETER, AS USED FOR STRENGTH OF BACK.

From D. Sargent, *Anthropometric Apparatus.*

NOTE.—The adjustment of the chain may, with advantage, be based upon *S*'s height. For this purpose, *E* may work out an empirical table of relations between height and the length of chain necessary to give the required position.

REFERENCES

(1) A. Binet and N. Vaschide, Expériences de force musculaire et de fond chez les jeunes garçons, in A. P., 4: 1897 (1898), 15–63.

(2) A. Binet and N. Vaschide, La mesure de la force musculaire chez les jeunes gens, *ibid.*, 173–199.

(3) W. Hastings. A manual for physical measurements for use in normal schools, public and preparatory schools, etc., Springfield. Mass., 1902. Pp. 112.

(4) D. Sargent, Strength tests and strong men at Harvard, in J. Boston Soc. Med. Science., No. 13, 1896-7.

(5) D. Sargent, Anthropometric apparatus, etc., Cambridge, Mass., 1887. Pp. 16.

(6) J. Seaver, Anthropometry and physical examination, New Haven, 1890.

TEST 8

Strength of legs.—This strength test is to be used in conjunction with strength of grip and strength of back. The best records in each of these three tests may be added, to secure an index of general bodily strength.

Instrument.—Back and leg dynamometer.

Method.—*S* stands upon the footrest of the instrument with his trunk and head erect and his chest well thrown out, but with the knees well bent (Fig. 12). *E* then adjusts the instrument so that the handle, when grasped by *S*, rests against his thighs. *S* should then take a full breath and give a hard lift, mostly with the legs, using the hands to hold the handle in place. Allow two or three trials as before.

Results.—Strength of legs is commonly about 26 per cent greater than strength of back. Thus, the 5000 men whose records are embodied in Hastings' table have a mean strength of legs of 189.5 kg., *P. E.*, 35.3, with a minimal record of 102.2 kg. and a maximal record of 276.8 kg.

TEST 9

Endurance of grip.—The object is to test the capacity of *S* to exert maximal muscular exertion, not in a single effort, as in Tests 6, 7, and 8, but during a period of one minute: the test is thus virtually identical with the endurance tests commonly undertaken by means of the ergograph.

Since Mosso's studies of muscular fatigue (31), the ergograph has been extensively employed, not only by physiologists, but also by psychologists and by investigators of school children. The form of the apparatus and the conditions of the test have been widely varied, and the numerous factors which affect the test have been exhaustively discussed. In general, the purposes for which the ergograph test has been employed may be summarized thus: (1) to study the

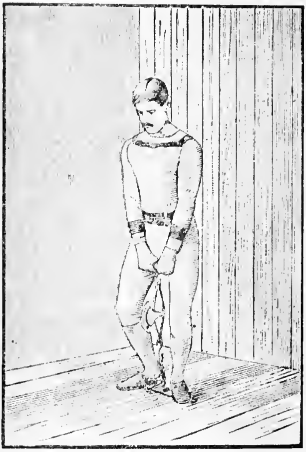


FIG. 12. BACK AND LEG DYNAMOMETER, AS USED FOR STRENGTH OF LEGS. From D. Sargent, *Anthropometric Apparatus.*

physiology of muscular contraction, (2) to detect the presence and to examine the nature and extent of muscular fatigue, (3) like the strength tests, to gain an index of physical capacity or endurance under varying conditions, *e.g.*, as affected by stimulants, narcotics, poisons, exercise, varying diets, etc., (4) to secure an index of right-handedness, (5) to discover whether physical fatigue is general or local, (6) to discover how mental work affects physical capacity, and, in particular, whether mental fatigue is reflected in muscular fatigue

with such clearness that its existence and degree may be ascertained by the examination of some restricted group of the muscles, and (7), on the assumption that physical capacity does measure directly the condition of mental efficiency, to determine the so-called diurnal 'course of power.'

As just intimated, the question of the applicability of the ergograph to these varied purposes raises a large number of problems, in particular that of the nature of fatigue. Since mental fatigue is reducible to physiological fatigue, the main point at issue is that of the precise nature of the latter.

The relative fatiguability of the structures concerned in physiological fatigue, *i. e.* of muscle, nerve fiber, and nerve cells, is a complex and disputed question. We know that the fiber, in comparison with other tissues, is extremely resistant to fatigue, either because the catabolic changes are minute, or, more probably, because they are at once compensated by adequate anabolism.

It has long been thought that the brain and spinal cord are much more susceptible to fatigue than the muscle, and consequently, the non-fatiguability of the nerve fiber has led most adherents of the neurone theory to look for the locus of nerve fatigue in the cell-body, in which, as has been shown by Hodge and others, excessive fatigue is accompanied by marked histologic changes.

On the other hand, Sherrington denies central fatigue to the cell-bodies, and locates it in the synapse (point where one neurone comes into functional relation with the next in the series), and assumes that this acts, like the motor end-plate (point where the neurone comes into functional relation with the muscle fiber), as a sort of safety-fuse to prevent overwork and damage to the tissues. Sherrington's experiment indicates that if a single, afferent tract is stimulated through several afferent tracts, the muscular contraction can be continued, when one afferent tract is exhausted, by recourse to the others. Other experiments, particularly those of Mlle. Joteyko, indicate that, compared with the terminal organs, the reflex mechanism of the spinal cord is practically indefatigable. If we extend this conclusion to the higher centers of the brain, we arrive at a peripheral theory of fatigue. Prominent evidence against such a conclusion has long been found in the observation that a muscle exhausted by volitional effort will still respond to electrical stimulation, but this observation is now discredited by the experiments of Kraepelin, G. E. Müller (32), R. Müller (33), Storey (38), and others, who have shown that the electrical stimulation in this case really sets in play a different set of muscles, and that, if a single muscle, like the *abductor indicis*, be properly isolated, the recuperation is not present.

"In view of these results and others," says Lee (25) "I am inclined to the belief that when we perform continued muscular work, our muscular system

fatigues before our central nervous system. Moreover, the same results make it probable that the brain and the spinal cord are, like the nerve fiber, resistant, and they throw a certain measure of doubt on all supposed proofs of central fatigue" (25, pp. 179-180).

The question of mental fatigue, induced by intellectual work, thus becomes perplexing. That it is a reality cannot, of course, be denied. Very likely even this type of fatigue is largely peripheral in origin, but how much is peripheral and how much central cannot as yet be stated. The usual assumption that brain fatigue is local in character seems due in part to the fact that the presence of fatigue is first indicated usually by the characteristic *feeling* of fatigue, but we may conclude that this feeling, like the corresponding feeling of effort, is primarily of peripheral origin, the consequence of the depressant action of certain toxic products, such as sarco-lactic acid, carbon dioxid, etc, which affect the muscular system in particular.

Since these toxic products, at least when present in considerable amounts, are carried by the blood throughout the body, it follows that decided fatigue is not confined to the tissues in which it arises (as Mosso showed in his experimental transfusion of the blood of a fatigued, into the veins of a fresh dog). Those who argue that fatigue is not general, but local, at least when not excessive, usually base their conviction upon the evidence of certain experimental studies which seem to indicate that well-defined local fatigue may exist without lowering perceptibly the functional capacity of other portions of the organism.

It may not be incorrect to assume that the toxic products are most detrimental in the locality where they are produced, yet may have, in time, a distinct general influence, and, similarly, that the reflex effects of fatigue, whether exciting or depressing, are partly local and partly general: the consumption of cell material, on the other hand, is more distinctly a local phenomenon (Bergström, 2).

In understanding the nature of mental fatigue and in distinguishing between general and local fatigue, it is helpful to separate the objective fatigue (*Ermüdung*), *i. e.*, actual functional inefficiency, from 'weariness' (*Müdigkeit*) *i. e.*, the subjective experience of ennui, loss of interest, or disinclination to work.[1] Thus, it is weariness rather than fatigue which disappears when one's occupation is changed: weariness is fluctuating, uncertain, and largely dependent upon the general conditions under which work

[1] Again some physiologists would differentiate fatigue and exhaustion: fatigue is, for them, a merely temporary reduction of efficiency, due to the accumulation of waste products, while exhaustion is a more serious condition due to lack of adequate nutrition. For a discussion of the nature of fatigue and of other factors such as practise, ennui, warming-up, spurt, etc., that complicate the determination of the fatigue curve, consult Kemsies (21), Kraepelin (22, 23), Thorndike (40) Ellis and Shipe (14), Bergström (2), Müller (33), Loeb (27), Aars and Larguier (1), Hough (19), Bettman (3), Rivers and Kraepelin (35), Weygandt (41), Lindley (26). An authoritative summary of the problem from the point of view of the physiologist will be found in Lee (25).

is being done, while fatigue increases more or less steadily and progressively during our waking moments. If weariness is often specific, fatigue may be more often general and operative to reduce the available energy for work (*Leistungsfähigkeit*) in any direction.

These considerations make it evident that, while the relations between mental fatigue and muscular energy are still obscure, we may hope, in principle, to secure some index of the former by our measurements of the latter. If the ergograph is to be employed for

FIG. 13. STOP-WATCH.

exact laboratory experimentation, there is no doubt but that the instrument must be of elaborate construction and the technique equally refined. The development of the ergograph itself from the relatively simple instrument of Mosso to such a complicated apparatus as that devised by Bergström is symptomatic of this gradual refinement of method and progressive analysis of the modifying factors.[1]

[1] Consult also Cattell (11), Franz (15), Binet and Vaschide (5), Binet and Henri (4), Bolton and Miller (10), Hirschlaff (17), R. Müller (33).

But while such problems as the merits of spring vs. weight loading, the relative capacity of a muscle working by isometric and by isotonic contractions, or the most reliable method of isolating the working muscle may be of paramount importance for laboratory investigation, it does not seem, on this account, absolutely impossible, as some writers assert, to secure valuable results for comparative purposes from large numbers of subjects by the use of simpler apparatus and less rigorous technique,—provided, of course, that

FIG. 14. METRONOME, WITH MERCURY CUPS FOR ELECTRIC CONTACT.

the conditions of experimentation are kept as constant as possible for different subjects and for the same subjects at different times.[1] For this reason, the test which is here described is suggested as a practical substitute for the more cumbersome and complicated ergograph.

APPARATUS.—Smedley dynamometer (Fig. 9). Stop-watch (Fig. 13). Metronome (Fig. 14). [If desired, a kymograph, (Fig. 15)

[1] In general, it may be expected that minor variable errors will, in a sufficiently long series of tests, be distributed according to the law of chance. Possibly, some portion of the dispute concerning the value of ergograms arises from the fact that certain experimenters have worked upon large numbers of subjects, while others have contented themselves with curves obtained from a single individual.

with drum support (Fig. 16) and other accessories, and a Marey tambour (Fig. 17), or the Mosso ergograph (Fig. 18).]

Two methods are described: one calls for a single, continuous contraction, the other for a series of separate contractions.

A. WITH CONTINUOUS CONTRACTION

METHOD.—Set the metronome at 60, *i.e.*, so that it beats once per second. Adjust the dynamometer to *S*'s hand, as in Test 6. Move the friction or recording pointer of the instrument well over to the

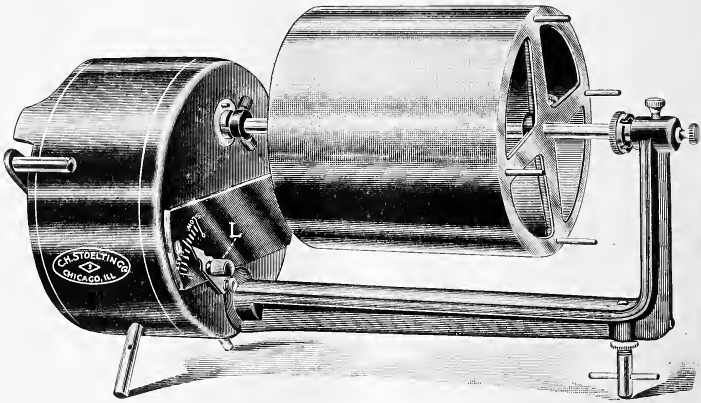


FIG. 15. KYMOGRAPH, IN HORIZONTAL POSITION.

A clock-work mechanism, with regulating fans, in the base, rotates the drum at constant speed and at any desired rate from one revolution in ten seconds to one revolution in ten minutes. Used for making graphic records upon smoked paper.

right, off the face of the scale. Instruct *S* that he is, at the signal 'now,' to grip as forcibly as possible, to maintain this grip with his utmost effort until told to stop at the end of one minute, and to keep his eyes fixed upon the pointer, so as to hold it as high as possible. (This instruction is designed to act as an incentive to maximal exer-

tion.) Let the instrument be held in the vertical plane with the right-hand edge resting on the table before which *S* is seated.

E starts the metronome, and, when he has caught the rhythm, starts the stop-watch, at the same instant saying 'now' for *S* to begin. *E* immediately takes the first reading, and thereafter glances at the scale at every fourth beat of the metronome. In the intervals, he records, of course, the reading of the pointer just obtained, estimating to the nearest half-kilogram. If the first reading is secured promptly, *E* will have 16 readings at the end of one minute.

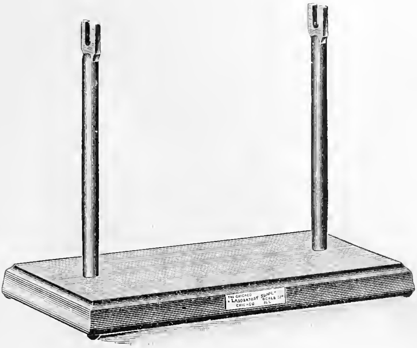


FIG. 16. SUPPORT FOR KYMOGRAPH DRUM WHILE BLACKENING PAPER.

VARIATIONS OF METHOD.—With the aid of the kymograph described in Test 10, *E* may secure a graphic record of *S*'s work, either by the use of the pneumatic tambour (the Smedley instrument is fitted for pneumatic transmission), or by the use of a simple system of levers to magnify the movement of the handle.[1] The quantitative evaluation of the resulting curve may then be obtained by a series of measurements of the ordinates taken at regular distances and checked by the record obtained as prescribed, or by

[1] For a cut showing the method of securing a dynamograph record, see MacDonald (29, p. 1184).

measurement with a planimeter of the area enclosed by the curve and its base line.

TREATMENT OF RESULTS.—(1) For some purposes, the results may be treated by simply averaging the 16 readings, but (2) it will usually be more instructive also to compare the initial with the final stages, in order to secure an index of endurance, or conversely, of fatigue. For this purpose, average the first four readings and the last four readings; subtract the latter from the former, and divide the remainder by the average of the first four readings. This may be expressed by the formula $x = \frac{r_1 - r_2}{r_1}$ when x = the desired index in terms of per cent, r_1 = the M of the first, and r_2 the M of the last readings. Or, (3) more simply, one may indicate endurance by the relation of the average to the maximal grip. (4) Following Binet and Vaschide (7), the records of strong, average, and weak S's (judged by their maximal grip) may be collated and treated in three groups, in order to trace the presence of the three types of endurance (see below).

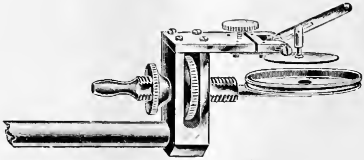


FIG. 17. MAREY TAMBOUR.

For securing tracings by pneumatic transmission. The rubber membrane is not shown.

B. WITH SEPARATE CONTRACTIONS

METHOD.—Adjust the metronome, dynamometer handle and pointer as in the first method. Inform S that, as the word is given, he is to make a series of 16 grips, each as forcibly as possible, and that these grips will be signalled at 4-sec. intervals. E then signals 'now' on every fourth beat of the metronome, and takes the readings as previously described.

VARIATIONS OF METHOD.—Substitute the kymograph tracing as suggested above. If this is done, there is no reason why the rate of effort may not be increased, so as to secure 60 or 120 contractions per minute, with a correspondingly more rapid onset of fatigue.

If it is desired to compare results obtained with the dynamometer with those obtained by the common form of ergographic experiment, it is suggested that E repeat the experiments made upon Chicago school children. For this purpose, substitute the Mosso ergograph for the dynamometer; use the kymograph for securing the graphic record, and the record furnished by the endless tape, multi-

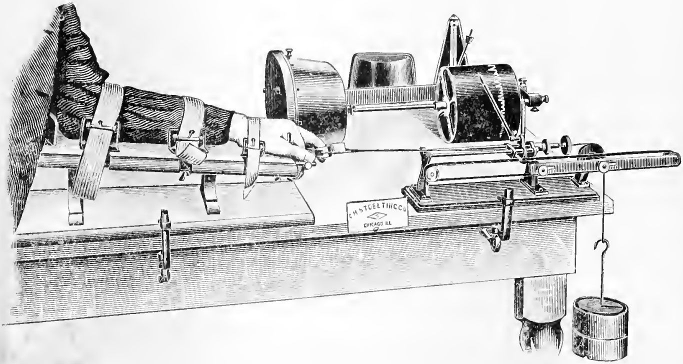


FIG. 18. MOSSO ERGOGRAPH, MODIFIED BY LOMBARD.

plied by the weight, for the quantitative result. Adjust the weight at 7 per cent of S's weight, and time the contractions to accord with the beats of a metronome set at 30, so as to secure 45 lifts in 90 sec.

TREATMENT OF RESULTS.—This may follow the lines already prescribed.

RESULTS.—(1) The measurement of endurance by the use of the dynamometer has been tried by Binet and Vaschide, though under conditions somewhat dissimilar to those we have suggested, upon a group of boys aged 10-13 years (6) and upon a group of young men aged about 18 (7). When five (or ten) grips with each hand, alter-

nately, were required, these authors made out four *types of endurance curve*, viz: (*a*) a sudden drop, then fairly constant, (*b*) an approximately stationary or constant type, which is quite common, (*c*) a continuous, but gradual drop, and (*d*) a more or less definite rise. The last is rather infrequent (it was not found, *e.g.*, by Clavière, in tests with 15 successive grips), but is sometimes given by vigorous individuals, though the third type is more common for such subjects. Practically 90 per cent of endurance records can, in the judgment of these writers, be classed in one of these four categories. Table 21 gives average records of five grips, made with the right hand, by groups representing these four types of endurance curve.

TABLE 21

Types of Endurance in Dynamometer Trials: kilograms (Binet and Vaschide)

	TYPE *a*	TYPE *b*	TYPE *c*	TYPE *d*
1st Grip........	23.00	18.70	24.12	17.33
2d Grip........	18.45	18.60	22.50	17.70
3d Grip........	19.00	19.20	21.17	18.67
4th Grip........	18.60	19.40	21.33	18.67
5th Grip........	18.20	17.80	19.80	20.67

(2) If we accept this hypothesis of 'types,' it is clear that the dynamometer yields a more reliable indication of the comparative muscular capacity when it is employed to test endurance in this way than when merely a single grip is taken, as in Test 6. To take an extreme, though actual case cited by Binet and Vaschide, it will be seen (Table 22) that, if two subjects belong to opposing types, their

TABLE 22

Opposed Types of Endurance, 10 Readings (Binet and Vaschide)

NUMBER OF GRIP	1	2	3	4	5	6	7	8	9	10
Subject B....................	36	34	34	30	29	28	25	26	26	29
Subject R....................	36	37	42	43	45	42	42	45	46	45

actual capacities may be completely unsuspected when but a single test is taken.

If we turn to the use of the ergograph, we find the following important, though too often conflicting results.

(3) Ergograph curves are affected by *practise* improvement, which, according to Bolton and Miller (10), results (*a*) from 'inurement,' *i.e.*, a fairly rapid "process of hardening and toughening of the skin where it comes in contact with the apparatus and of habituating the muscles to the strains which the unusual effort imposes," (*b*) from improved coördination in the movements concerned, particularly seen in the disappearance of useless movements, (*c*) from improvement in the rhythmic execution of the contraction, and (*d*) from a slow increase in endurance proper, primarily in the nerve centers. This increase of practise, as Oseretzkowsky and Kraepelin (34) have shown, affects both the height and the number of lifts, and gradually becomes less and less noticeable as maximal practise is attained.

(4) The amount of work that can be done by the muscle is increased if the *rate of lifting* is increased from 30 to 60 or 120 lifts per minute (Oseretzkowsky and Kraepelin).

(5) The work done is conditioned by the *load lifted* or tension of the spring. One can not, without caution, compare ergograms made with different loads.

(6) The *total amount of physical work* done, as measured by weight × distance, can not be regarded as a necessarily correct index of the physiological capacity of the muscle; thus, 100 lifts of 25 mm. each may not be assumed to be physiologically equal to 50 lifts of 50 mm. each (Binet and Henri, 4; Franz, 15).

(7) The *weight ergograph* is not adapted to the measurement of muscular capacity (Binet and Henri, 4), hence "the fatigue curves obtained by Mosso and later investigators with weights do not represent the true state of the neuro-muscular system" (Franz, 15).

(8) "The *isotonic* use of a weight or a spring for measuring muscular force is not justified, because two variable factors, extent and force, are introduced," so that an isometric spring (such as the dynamometer) should be used for all comparative experiments (Franz).

(9) With improperly contrived apparatus or inexperienced subjects, the ergographic tracing is very liable to be affected by the play

of *muscles* other than those under examination (Binet and Henri, 4. Bergström, 2). Müller (33) considers the failure properly to isolate the muscle a fundamental defect of the ergograph.

(10) In addition to these specific criticisms, more general *conclusions of a negative character* may be quoted. Thus, Bolton (8) asserts that the ergograph is not adapted for measuring the degree of fatigue in school children; Bolton and Miller (10) conclude that ergograph records "have slight validity until inurement has become thorough and coördination complete, that the ergograph is quite unadapted to the obtaining of exact statistics upon a large number of individuals, and that records taken upon unpractised subjects, both before and after operations whose influences are thought to affect muscular power, are without the slightest claim to trustworthiness." Similar conclusions are reached by Ellis and Shipe (14), after a retrial of the methods of Keller and of Smedley also by Thorndike (40) and, with some qualifications, by Bergström (2).

(11) The *effect of physical work* upon ergographic curves seems to vary with the physical condition of the individual and with the nature and duration of the exercise. Thus, Bolton (8, 9) found his ergograms decreased by a 2-hour walk, but Oseretzkowsky and Kraepelin found that a 1-hour walk caused at first a transient improvement, then a reduction, the first of which they attribute to the increased excitement of central motor tracts, and the second to the dampening influence of general muscular fatigue. Smedley (37) tested Chicago children before and after a 40-minute class exercise in the gymnasium with the result that the stronger pupils were little affected, whereas weak and nervous pupils were decidedly exhausted. From this study he concluded that the classes in physical culture should be graded on a physical, instead of on an intellectual basis.

(12) Extensive study of the *effect of mental work* on physical endurance has so far yielded but discordant results. Some of this work, *e.g.*, that of Keller (20), may be thrown out of court at once as careless in plan and execution and merely illustrative of blind infatuation for the ergograph. Typical conclusions of other investigators are as follows: Larguier (24) reports that two hours of mathematics, and Bolton that two hours of adding, definitely increase the

ergograph record; Clavière (13), on the other hand, reports that two hours of intense mental work produces a definite and proportionate diminution of muscular force, whereas intellectual work of medium intensity does not produce any appreciable weakening of endurance; he further confesses his inability to determine the relative fatigue-effect of various school studies. The careful ergographic tests of Oseretzkowsky and Kraepelin show that work-capacity is increased after one hour of simple addition or learning of 12-place numerals, but that it is lessened if the mental work is rendered more difficult, as by adding under distraction. In an extensively quoted study, Kemsies (21) reports the results of a long series of tests upon a selected group of average, industrious boys who had been trained to the use of the instrument, from which he concludes (*a*) that the ergograph is a reliable indicator of true fatigue (lowered fund of energy as distinct from weariness), (*b*) that subjective feelings of bodily or mental condition may not accord with real capacity, (*c*) that some of the pupils in the Berlin schools show, at least for the time being, signs of overwork, (*d*) that special attention should be paid to pupils who fatigue easily, (*e*) that one can determine for each study its special fatigue-value or 'ergographic-index,' more particularly, that the several studies range themselves, in order from highest to lowest fatigue-index, as follows: gymnastics, mathematics, foreign languages, religion, German, science and geography, history, singing and drawing.

In the attempt to explain these divergences, Binet and Henri (4) suggest that we must always distinguish between mental work conducted without emotion and that conducted with emotion; they conclude that the former, if prolonged, may be expected to lessen endurance, the latter to produce a transient increase followed by a decrease. Kraepelin (23), somewhat similarly, concludes that, while hard mental work certainly reduces muscular energy, deviating results may appear in ergograms on account of the condition of excitement (*Anregung*) that normally accompanies mental work, and that may be expected to affect, either positively or negatively, the tracing which follows such work. Kraepelin further calls attention, as do Ellis and Shipe, Bergström, Franz, and others, to the very large normal variation in the curves of any individual, due to the operation of numerous constant and variable fac-

tors, often little understood. Many results are valueless (*e. g.*, in his opinion, those of Kemsies) because of the failure properly to eliminate or evaluate these factors.

(13) The investigations of Christopher (12) and Smedley (37) at Chicago indicate a thorough-going correlation between *endurance and class-standing*, according to the method of percentile grading, the method of distribution of 12-year old pupils, and the method of comparison of the endurance of children at and above grade with that of children below grade at each age. Again, boys in the school for incorrigible and truant children were found to exhibit, at every age, less endurance (62 per cent to 82 per cent) than normal boys of the same age.

(14) The *endurance of boys* is greater than that of girls at all ages, and the difference becomes very striking during adolescence (37).

(15) The *development of endurance and that of vital capacity* bear a decided resemblance to one another, whether pupils are examined singly or collectively (37).

(16) The *diurnal 'course of power'* according to the Chicago experiments may be expressed as follows: "(*a*) The extremes of endurance and fatigue in school are greater in the morning than in the afternoon; (*b*) a higher grade of power is found in the morning session in children attending two sessions daily; (*c*) while endurance is not as great, it is better sustained in the afternoon." Compilations of the ergograms of 1127 pupils place the maximum at 9 a.m. and the minimum at 12 noon. Kemsies considered the first two morning hours the best. Experiments upon adults by Lombard (28), Harley (16), Storey (39), and Marsh (30) exhibit considerable lack of agreement with one another or with the Chicago results, though Marsh summarizes them by the statement that the curve of strength efficiency seems well established for the following course: "a beginning minimum in early morning, a fairly rapid rise till 11, a level or slight decline till 1 p.m. (± 1 hour), an increase to the maximum at 5 p.m. (± 1 hour), thence a fall till bed-time."

(17) Kemsies concludes that Monday and Tuesday, or the first two days after any rest pause, are the *best days for general efficiency*, and he further concludes that vacations exert a powerful effect upon efficiency, but, since this effect can not be traced for longer than four weeks, school terms should be broken up by more frequent vacations of shorter duration.

(18) If ergographic contractions are continued to the point of *exhaustion*, we have both the sum total of the height of the lifts and their number for indexes of the neuro-muscular condition. Hoch and Kraepelin (18) are of the opinion that, in this case, the height of contraction is conditioned by the state of the muscles, but the number of contractions by the state of the central nervous system; the two factors should, therefore, be reported separately for their diagnostic value. On the other hand, Lombard (28) concludes that, at least when the contraction is not faster than once per second, the amount of fatigue experienced by the central nervous system does not correspond to the number of lifts, but rather to the strength of the motor impulses discharged, so that the sum total of the height of lifts is the more accurate index of the state of the central nervous mechanism.

(19) According to Lombard, endurance is increased by exercise, rest (especially sleep), food, increased atmospheric pressure, and by small doses of alcohol, but lessened by general and local fatigue, hunger, lessened atmospheric pressure, high temperature, especially with high humidity, and by tobacco. Oseretzkowsky and Kraepelin find that coffee increases the height of lifts, and that alcohol, in quantities from 15 to 20 g., causes at first a considerable increase, especially in the number of lifts, but that this soon disappears. On the other hand, Rivers and Webber (36) have discovered that small doses of alcohol (5-20 cc.) fail to produce any appreciable modification of the ergographic record if proper precautions are taken to keep the subject in ignorance as to when alcohol is administered. The results of previous workers are therefore presumably due to the influence of other factors, particularly interest and sensory stimulation, and no future work on the effects of small doses of alcohol can be acceptable unless these factors are controlled.

Harley (16) concludes that "moderate smoking, although it may have a slight influence in diminishing the power of doing voluntary muscular work, neither stops the morning rise nor, when done early in the evening, hinders the evening fall."

NOTES.—In either form of test, E must practise his work until it becomes automatic. He must take care to keep his eyes directly over the pointer to prevent the error of parallax in reading. For this reason it will be found most convenient for S and E to sit on opposite sides of the table.

The stop-watch is used both to test the metronome and to check up the duration of the experiment, but ought to be virtually unnecessary after *E* has practised the experiment sufficiently.

In the first form of test, the pointer is apt to fall by a series of sudden drops, or even at times to rise as *S* makes a momentary recovery. *E* must take the reading precisely on the beat of the metronome, regardless of the position of the pointer just before or just after the beat.

To hasten the acquisition of skill in conducting the second form of test, *E* will find it helpful to accent the spoken 'now' and to get the swing of the four-beat rhythm by mentally counting the other beats, thus: "*Now*, two, three, four: *now*, two, three, four." As soon as the utterance becomes automatic, *E* can give his whole attention to the readings and the recording of them, and an accurate record can be obtained from very quick and brief excursions of the pointer. Incidentally, some *S*'s may be found who are inclined to hold the pointer up too long: they must be cautioned against this, otherwise fatigue will set in very rapidly.

REFERENCES

(1) K. Aars and J. Larguier, L'effort musculaire et la fatigue des centres nerveux, in A. P., 7: 1900 (1901), 187-205.

(2) J. A. Bergström, A new type of ergograph, with a discussion of ergographic experimentation, in A. J. P., 14: 1903, 510–540.

(3) S. Bettmann, Ueber die Beeinflussung einfacher psychischer Vorgänge durch körperliche u. geistige Arbeit, in P. A., 1: 1895, 152-208.

(4) A. Binet and V. Henri, La fatigue intellectuelle, Paris, 1898. Pp. 336.

(5) A. Binet and N. Vaschide, Examen critique de l'ergographe de Mosso, in A. P., 1897 (1898), 253-266. Also, Un nouvel ergographe, dit ergographe à ressort, *ibid.*, 303-315.

(6) A. Binet and N. Vaschide, Expériences de force musculaire et de fond chez les jeunes garçons, *ibid.*, 15-63.

(7) A. Binet and N. Vaschide, La mesure de la force musculaire chez les jeunes gens, *ibid.*, 173-199. See also *ibid.*, 236-244, 245-252, and 295-302.

(8) T. Bolton, The reliability of certain methods for measuring the degree of fatigue in school children, in P. R., 7: 1900, 136-7.

(9) T. Bolton, Ueber die Beziehungen zwischen Ermüdung, Raumsinn der Haut und Muskelleistung, in P. A., 4: 1902, 175-234.

(10) T. Bolton and Eleanora Miller, On the validity of the ergograph as a measurer of work capacity, in Nebraska Univ. Studies, 1904, 79 + 128.

(11) J. Cattell, An ergometer, in Science, n.s., 5: 1897, 909–910.

(12) W. Christopher, Rept. on child-study, reprint from An. Rept. Brd. Educ. Chicago, 1898–99.

(13) J. Clavière, Le travail intellectuel dans ses rapports avec la force musculaire mesurée au dynamomètre, in A. P., 7: 1900 (1901), 206-230.

(14) A. C. Ellis and Maud Shipe, A study of the accuracy of the present methods of testing fatigue, in A. J. P., 14: 1903, 496-509.

(15) S. I. Franz, On the methods of estimating the force of voluntary muscular contraction and on fatigue, in Amer. J. Physiol., 4: 1900, 348-372.

(16) V. Harley, The value of sugar and the effect of smoking on muscular work, in J. of Physiol., 16: 1894, 97-122.

(17) L. Hirschlaff, Zur Methode und Kritik der Ergographenmessungen, in Z.P.P., 3: 1901, 184-198.

(18) A. Hoch and E. Kraepelin, Ueber die Wirkung der Theebestandtheile auf körperliche und geistige Arbeit, in P. A., 1: 1896, 378–488.

(19) T. Hough, Ergographic studies in neuro-muscular fatigue, in Amer. J. Physiol., 5: 1901, 240-266.

(20) R. Keller, Pädagogisch-psychometrische Studien, in Biol. Centralblatt, 14: 1894, 24-32, 38-53, 328–336.

(21) F. Kemsies, Zur Frage der Ueberbürdung unserer Schuljugend, in Deutsche med. Wochenschrift, July 2, 1896, 433. See also his Arbeitshygiene der Schule auf Grund von Ermüdungsmessungen, Berlin, 1898, and in S. Z., 2: 1899, Heft i., Pp. 64.

(22) E. Kraepelin, Zur Ueberbürdungsfrage, Jena, 1897. Pp. 49.

(23) E. Kraepelin, Ueber Ermüdungsmessungen, in A. G. P., 1: 1903, 9–30.

(24) J. Larguier, Essai de comparison sur les différentes méthodes proposées pour la mesure de la fatigue intellectuelle, in A. P., 5: 1898 (1899) 190-201.

(25) F. S. Lee, Fatigue, in the Harvey Lectures, Phil., 1906, 169-194. Also in J. Amer. Med. Ass., 46: 1906 1491, and in Studies in Physiology, Columbia Univ., 1902–7.

(26) E. H. Lindley, Ueber Arbeit und Ruhe, in P. A., 3:1900, 482-534.

(27) J. Loeb, Muskelthätigkeit als Mass psychischer Thätigkeit, in Arch. f. d. ges. Physiol., 39: 1886, 592-7.

(28) W. P. Lombard, Some of the influences which affect the power of voluntary muscular contractions, in J. of Physiol., 13: 1892, 1–58.

(29) A. MacDonald, Experimental study of school childern, etc., reprint of chs. 21 and 25 of U. S., 1899.

(30) H. D. Marsh, The diurnal course of efficiency, Columbia Univ. diss., N. Y., 1906. Pp. 99.

(31) A. Mosso, Fatigue, Eng. tr., N. Y., 1904. Pp. 334.

(32) G. E. Müller, Review of A. Waller's "The Sense of Effort," in Z. P., 4: 1893, 122-138.

(33) R. Müller, Ueber Mosso's Ergographen, etc., in Ph. S., 17: 1901, 1-29.

(34) A. Oseretzkowsky and E. Kraepelin, Ueber die Beeinflussung der

Muskelleistung durch verschiedene Arbeitsbedingungen, in P. A., 3: 1901, 587-690.

(35) W. Rivers and E. Kraepelin, Ueber Ermüdung u. Erholung, in P. A., 1: 1896, 627-678.

(36) W. Rivers, The influence of small doses of alcohol on muscular activity, in P. B., 5: 1908, 49. See also Rivers and H. Webber, The influence of small doses of alcohol on the capacity for muscular work, in B. J. P., 2: 1908, 261-280.

(37) F. Smedley, Rept. dept. child-study and pedagogic investigation, reprint from 46th An. Rept. Brd. Educ., Chicago, 1899-1900. Also in U. S., 1902, i., 1095-1115.

(38) T. Storey, The influence of fatigue upon the speed of voluntary contraction of human muscle, in Amer. J. Physiol., 8: 1903, 355.

(39) T. Storey, (*a*) Some daily variations in height, weight, and strength, in Amer. Phys. Educ. Rev., 6: 1901. (*b*) Daily variation in the power of voluntary muscular contraction, *ibid.*, 7: 1902. (*c*) Studies in voluntary muscular contraction, Stanford Univ. Press, 1904.

(40) E. L. Thorndike, Mental fatigue, in P. R., 7: 1900, 466–482, 547–579.

(41) W. Weygandt, Ueber den Einfluss des Arbeitswechsels auf fortlaufende geistige Arbeit, in P. A., 2: 1897, 118–202.

TEST 10

Quickness or rate of movement: Tapping.—This has probably been more frequently applied than any other 'motor test,' and has been thought to afford a better index of motor capacity than any other single test. Recent work with tapping, however, while not discouraging the belief that the test has value, has shown that we cannot regard speed of voluntary movement as an unequivocal and comprehensive 'index of voluntary motor ability,' because a high gross rate does not necessarily go hand in hand with high speed in other phases of motor response, and because, moreover, we do not know precisely what may be the ultimate neural or psychophysical factors that condition the rate.

Aside from its use in the attempt to secure this 'index of voluntary motor ability,' the tapping test has been employed to secure an index of righthandedness (for which purpose it may be advantageously combined with Tests 6, 9, 11, and 12), and to secure an index of fatigue (likewise preferably in conjunction with other tests of physical capacity). These several indexes have been studied in various comparative investigations, more especially in estimating

sex and age factors in motor development and the relation of physical to mental ability at large.

The method has ranged from the very simple making of dots or vertical marks with pencil and paper (Binet and Vaschide) to the execution of difficult trilling movements upon telegraph keys. The apparatus here prescribed is somewhat elaborate, but experience has shown that the tapping test cannot be conducted without careful control of experimental conditions, and the use of a reliable recording device, such as the graphic method supplies.

Tests like rapid counting aloud or the rapid reading of digits or the reaction-time test, are not psychologically comparable to the tapping test. Again, the form of test used at Columbia University and elsewhere to measure rate of movement (making a dot as rapidly as possible in each of 100 one-cm. squares) is not equivalent to the tapping required in most quickness tests, since a certain

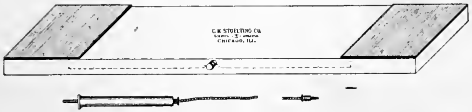


FIG. 19. TAPPING-BOARD.

amount of precision is demanded of each movement, and that test therefore stands midway between Tests 10 and 11, as here prescribed.

MATERIALS.—Tapping board, 55 x 10 cm., with brass plates 10 cm. square on either end (Fig. 19). Tapping stylus, with flexible connecting wire attached. Kymograph (Fig. 15) with accessories,—paper, smoking device, shellac solution. Double time-marker (Fig. 20). Seconds' pendulum (Fig. 21) or other noiseless instrument arranged to give electric contacts once per sec. Support with levelling screw and right-angle piece to hold time-marker. Table clamps for tapping board. Large sheet of gray or white cardboard. Suitable supports and clamps for holding cardboard. Two short-circuiting keys (Fig. 22), or simple knife switches. Stop-watch. Four dry or Leclanché cells. Flexible covered wire with connector tips or ordinary No. 18 annunciator wire. [A swivel

chair adjustable in height and an ammeter or battery-tester are also convenient, though not absolutely essential.]

Preliminaries.—(1) Clamp or screw the tapping board securely to the side of a table in such a manner that *S* may have free access to either end of the board for using either right or left hand. Arrange *S*'s chair so that he sits sidewise to the table with his forearm resting comfortably along the tapping board and his hand directly over the metallic plate.

(2) Place the kymograph in a horizontal position, screened from *S*'s view by the sheet of cardboard.[1] Adjust the fans or gear-

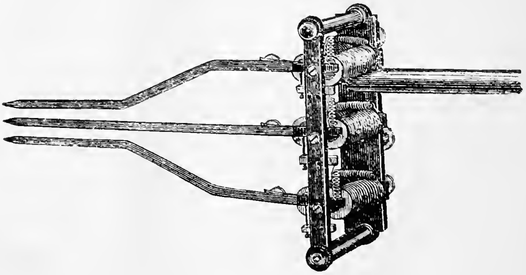


FIG. 20. TRIPLE TIME-MARKER.

The double and the single time-marker are of similar construction.

wheels so that the drum makes (for a 30-sec. test) one revolution in about 40 sec.

(3) Remove the drum and cover with the prepared paper by simply moistening the gummed end, taking care to draw it evenly and tightly around the drum. Blacken the paper by revolving it slowly in a smoky flame.[2] Replace carefully in the kymograph.

[1] The screen is to avoid the distraction of *S*'s attention by the operation of the apparatus. If separate tables are used for tapping board and kymograph, this may not be necessary, but it is commonly more convenient to assemble all the apparatus on a single table.

[2] An oil stove from which the top is removed is excellent for this purpose, as the flame is very sooty and not so hot as the gas flame often employed. A simple support (Fig. 16) is used to hold the drum, both for the smoking and for the subsequent removal of the paper. For this and other details in the

(4) Adjust the time-marker on the support so that the pointers bear upon the drum with just sufficient pressure to make a satisfactory tracing. The pointer must move in a plane parallel to the plane of a tangent drawn through the point of contact.

The manipulation of the apparatus may be facilitated by fastening upon the table, in front of, and parallel to the surface of the drum, a straight bit of wood somewhat longer than the drum. Let the foot of the tripod which contains the levelling screw stand away from the drum, and the other two feet bear against the wooden strip. A half turn of the levelling screw will then free the pointers from the drum, and the entire support with the time-marker may be slid along to a new position, when another half turn of the screw will quickly adjust the pointers for the next record.

(5) Wire one signal-magnet in series with the tapping board stylus, short-circuiting key, and two cells of the battery. The magnet will then be set in motion by the tapping when the key is closed.

(6) Wire the second magnet in series with the pendulum, second short-circuiting key, and remaining two cells of the battery. This magnet will then be set in motion by the pendulum when its controlling key is closed, and will thus beat off the time-line.

METHOD.—(1) Seat S for the use of his right hand. Instruct him to tap as rapidly as possible from the signal 'now' to the signal 'stop,' which will be given about one-half minute later. Tell him to pay attention only to his tapping. He may be allowed to exercise some latitude with regard to the type of movement used (short or wide excursion), unless he is inclined to adopt a very heavy whole-arm pounding movement. The most favorable movement for most S's is that obtained by resting the elbow on the tapping board and using both the wrist and elbow joints.

(2) Start the seconds' pendulum and close the time-line circuit.

use of the kymograph, consult Titchener, *Experimental Psychology*, vol. I, Part II, pp. 172–180.

If an extended series of tests is to be made, cover the drum permanently with the regular kymograph paper, and, for the records, superpose two narrower strips, say 75 mm. wide. These strips are wide enough to record the right and left-hand efficiency of one S: they can then be removed promptly for fixing, and thus the danger of injury is lessened, the ease of handling increased, and the blackening of the metal drum is less likely to be a source of annoyance.

(3) Start the kymograph, and at the same time give S the signal 'now.'

(4) When S is fairly started, throw in the record magnet by closing its key, and at the same instant start the stop-watch.

(5) At the expiration of 30 sec., break the record circuit, signal 'stop' to S, stop the kymograph and the watch, and open the time-circuit. [E must practise the whole series of operations until they

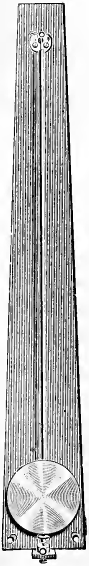


FIG. 21. SECONDS' PENDULUM.

run smoothly and automatically, especially the simultaneous operation of watch and record key.]

(6) Now adjust the pointers for a new record. Let S sit facing in the other direction, and test the left hand by the use of the plate at the other end of the tapping board, following the directions given for the test of the right hand.

VARIATIONS OF METHOD.—(1) The duration of the test may be lengthened to 45 sec. or longer, or shortened to 10 or 20 sec. It is, however, desirable for the sake of comparison that a standard duration be employed. Thirty sec. is adequate for all ordinary purposes.

(2) It is recommended that, whenever time permits, more than one trial be made for each hand. To follow the procedure suggested by Wells, *E* should first make five trials with the right, then five trials with the left hand. Each trial lasts 30 sec., and is followed by a rest-pause of 2.5 min., during which *S* should refrain from all muscular effort. The five trials of 30 sec. each constitute one 'record,' and the two records of right and left hand constitute one 'experiment.'

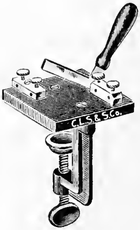


FIG. 22. SHORT-CIRCUITING KEY (DU BOIS REYMOND).

(3) To make the test comparable to the form employed by some investigators, an ordinary 'sending' telegraph key may be substituted for the tapping board and stylus. But the key has the disadvantage of imposing a certain restriction upon the type of movement, and will be found in practise to reduce the record of many *S*'s.

(4) By using the key, *E* may compare *S*'s rate of tapping with different fingers. For this purpose, it is well to fasten down the forearm with a strap at the wrist, so as to allow movement with the fingers only.

(5) Again, by using the key, a trilling movement, executed by alternate movements of the index and middle fingers, may be substituted for the regulation tapping movement. Without practise, this movement is quite difficult for some *S*'s, whereas for others,

notably for those who have practised trilling exercises on the piano, it is comparatively easy.[1] For this reason, this form of experiment is not advised, save for some exceptional purpose, *e.g.*, testing the effect of practise upon the acquisition of a new bit of manual dexterity. Other modifications suggest themselves, such as trilling with the 4th and 5th fingers—an exercise likely to be unfamiliar even to *S*'s who have 'taken lessons.'

TREATMENT OF RESULTS.—(1) When the record has been made, use any pointed article to mark it for future identification (*S*'s name or number, date, hand used, etc.); then remove carefully for preservation. A simple and satisfactory method is to pour a very thin solution of shellac and wood alcohol, or of powdered resin (not over 10 per cent) in alcohol, into a saucer or shallow dish, and to pass the strip through this, smoked side up. Hang the record up to dry, and pour the solution back into a wide-mouthed bottle, where it should be kept tightly stoppered. The record will dry in a few minutes and can then be trimmed and handled with impunity.

(2) The result of the test is commonly expressed simply by the total number of taps executed, but it is quite as important, if not more important, to consider changes of speed during each trial. The requisite data must be secured by the rather laborious process of counting the strokes made by the recording-magnet upon the blackened paper, and tabulating them by 5-sec. intervals, as illustrated below. The 'total efficiency' of a 'record' (5 trials with the same hand) is the average of the sum of the taps per trial. This serves as the gross index of speed: the rates for the 6 intervals within each trial afford an opportunity for studying variations in performance.

The use of an electric counter, such as some investigators have employed, would eliminate this work, but the counter gives no indication of changes in speed during the trial. Moreover, the electric counter is not reliable: even with 10 or 12 cells of battery, it will miss a quick tap which the graphic method will record. It follows that all results based on the use of the counter are to be looked upon with suspicion, so that the conclusions of Bagley,

---

[1] The effect of piano practise, as the investigations of Binet and Courtier (2) and of Raif (14) show, is to improve coördination of movement, *i, e.*, its regularity, smoothness, etc., but not to increase the natural capacity for speed or rate of movement.

Bolton, Marsh, Kelly, Smedley, and possibly those of Bryan, Davis, Gilbert, and Dresslar should be accepted with reservation.

(3) To secure an index of fatigue, E may compare the record of the first 5 (or 10) sec. with that of the last 5 (or 10) sec. by the use of the formula for determining the relative loss of efficiency given in Test 9.

Wells has published extensive conclusions concerning fatigue in tapping that are based upon a differently computed index. The average number of taps executed in the 2d, 3d, 4th, 5th and 6th 5-sec. intervals are divided by the number of taps executed in the 1st 5-sec. interval. This index is somewhat misleading, in so far as a high index indicates a low degree of fatigue. If it is deemed worth while to relate the last five to the first of the six intervals to compute fatigue, it would be better, in the author's opinion, to subtract the average in question from the initial speed and divide the loss by the efficiency of the first interval.

TABLE 23

Sample Record of a Tapping Test (Wells)

NUMBER OF INTERVAL	1ST	2D	3D	4TH	5TH	6TH	TOTAL
1st trial	41	37	35	34	34	32	213
2d trial	41	37	36	35	34	34	217
3d trial	40	39	37	37	35	34	222
4th trial	40	39	37	36	36	35	223
5th trial	41	39	38	37	36	36	227
Averages	40.6	38.2	36.6	35.8	35.0	34.2	220.4

(4) To secure an index of righthandedness, E may compute the percentage of the left-hand to the right-hand efficiency. The fatigue-index of the right hand may also be compared with that of the left hand in a similar manner.

Typical Results.—Table 23 shows a sample record of the work of a normal adult with the right hand, when near the limit of practise. The tabulation is in accordance with that recommended when five 30-sec. trials are made.

General Conclusions.—Although the tapping test is one of the most objective that can be applied, and although it has been tried by a large number of investigators (see the references at the end of

the test), the results have not been always accordant, and, with the exception of the recent work of Wells, have not been so treated as to afford real insight into the factors that underlie their appearance. The lack of accordance is to be attributed in large part to differences in method of procedure. Differences in apparatus, too, have been sufficient to account for some discrepancy, as has already been pointed out. As regards method, the duration of the test, to instance a single point, has varied from 5 sec. (Binet and Vaschide, Bryan, Kirkpatrick) to 2 min. (Thompson), with intermediate durations, such as 10 sec. (Bagley), 30 sec. (Smedley), 45 sec. (Gilbert), 60 sec. (Kelly), or the test has been conducted in 5 series of 5 sec. each (Bolton).[1]

In so far as these divergences of method may be neglected, we may note the chief conclusions of interest concerning the tapping test, as follows.

(1) In general, the *maximal rate* of voluntary movement varies with the individual, with sex, with maturity, with the side of the body used, with practise, with the number of trials (duration of experiment), with fatigue, with mental excitement, with the time of day, but not, within wide limits, with the amplitude of the movement.[2]

(2) *Constant individual differences* in rate of tapping can be demonstrated without much difficulty, but we cannot at present explain them, save to say that they are conditioned in a general way by fundamental neural factors, or by these plus differences in ability to coördinate voluntary movements.[3] Thus, in 10 adults tested by Wells (10 trials for each hand), the average total efficiency (taps in 30 sec.) was approximately 194, but the fastest *S* averaged 225,

[1] The situation here, as in most tests, shows clearly how desirable it would be to establish some standard form of test and to use it alone for all comparative purposes.

[2] For a fuller discussion of these conditions, consult Dresslar, Bryan, and Wells.

[3] To quote from Wells (19, p. 444): "What is the precise physiological significance of the maximum rate is by no means well made out. . . . It seems to be generally conceded that it is limited by the refractory phase of the synapses in the motor pathways, but that does not make the tapping test a measure of the period of this refractory phase; at least, not in the earlier stages of practise. . . . In the beginning, as we ordinarily have to apply the test, the factors in speed are probably those of coördination mainly, and cannot be expected to afford information about the condition of the motor pathways as given in the refractory phase."

and the slowest 153. Since the m.v. is small (here approximately 1 to 3 per cent) these figures undoubtedly indicate persistent characteristic differences. In general, it may be said that, for adults, initial right-hand rates range from 5 to 14 taps per sec.

(3) The rate of tapping increases with *age*, at least between 6 and 18 years. The slight drop at 13, upon which Gilbert comments, appears in Bryan's tables with some qualifications, but not so clearly in Smedley's results, which are reproduced herewith: it will be seen, however, that boys make no apparent gain from 13 to 14.

TABLE 24

Dependence of Rate of Tapping upon Age (Smedley)

AGE	BOYS: NUMBER TESTED	BOYS: TAPS IN 30 SECONDS, Rt. Hand	BOYS: TAPS IN 30 SECONDS, Lt. Hand	GIRLS: NUMBER TESTED	GIRLS: TAPS IN 30 SECONDS, Rt. Hand	GIRLS: TAPS IN 30 SECONDS, Lt. Hand
8..........	31	147	117	31	146	117
9..........	60	151	127	44	149	118
10..........	47	161	132	48	157	129
11..........	49	169	141	48	169	139
12..........	44	170	145	50	169	140
13..........	50	184	156	45	178	153
14..........	40	184	155	67	181	157
15..........	37	191	169	48	181	159
16..........	21	196	170	50	188	167
17..........	13	196	174	40	184	162
18..........	3	197	183	24	193	169

(4) *Sex.* The results of most investigators lead to the conclusion that boys are faster than girls, and that this sex difference increases with age. Bolton, however, has reported that "the girls are uniformly better than the boys," while Bryan found girls superior at 13, when they showed improvement and the boys little or none,—a tendency that is apparently allied to the actual crossing of the curves of height and weight. More extensive experiments upon adults (10 men and 10 women) by Wells (21) now indicate that women surpass men in tapping with the right hand in the first experiment, whereas elsewhere they are inferior: the sex differences

found by this investigator are said to be "mainly in those features of the experiment which especially involve the affective factor in the subject's attitude; and they are manifestations of the greater responsiveness of the women to this affective element."

(5) The *index of righthandedness* (per cent left-hand is of right-hand efficiency) was found by Wells to range from .81 to .94, average .90, for adults, and by Smedley to vary with age in the case of school children, in such a manner that the average index was .82 at the age of 9, and .89 at the age of 18. It is evident, therefore, that righthandedness, so far as tapping is concerned, is more pronounced in childhood than in adult life. Wells also states (21) that "the right and left hands are farther apart in women," though the relationship is more variable in them than in men.

(6) *Righthandedness and intelligence.* Smedley's conclusion that there exists a positive correlation between degree of righthandedness and school standing, *i. e.*, that the left-hand more nearly approaches the right-hand efficiency in the case of dull and backward pupils is not confirmed by the results of Bolton.

(7) *Warming-up.* In practically every continuous psychophysical activity there appears a tendency to improvement due to what the Germans have termed *Anregung*. This 'warming-up' is a kind of momentum, not identical with practise, and its effect is to increase or heighten the activity, and thus to retard or even to obscure the appearance of indications of fatigue. In tapping, we observe fatigue within each 30-sec. trial, but a comparison of successive trials within a record will show the improvement due to warming-up. With 2.5 min. rest-pauses, Wells found this factor to be clearly present (up to the 7th trial at least) in right-hand records, but by no means so evident in left-hand records. The effect of warming-up appears to be primari y operative in increa ed immunity to fatigue, and is markedly augmented by practise, *e.g.*, in tests continued for 20 days.

(8) *Spurts.* The curve of performance in tapping, as well as any psychophysical activity, is also liable to be influenced by short periods of increased activity, which, to continue the analogy of the race-track, may be termed 'spurts' (German, *Antriebe*). Thus, Wells' discovery that the first experiment usually excels the second in women whereas the reverse may be true in men, is re-

ferred to a special incitement of novelty (*Neuigkeitsantrieb*), which affects the women markedly Similarly, each 'record' may be affected by an initial spurt (*Anfangsantrieb*) or by a terminal spurt (*Schlussantrieb*). These dynamogenic factors obviously tend to obscure the real effects of fatigue.

(9) *Fatigue and the fatigue-index.* (*a*) As just stated, the speed of tapping normally declines after the 1st 5-sec. interval, until it is approximately $\frac{6}{7}$ as great in the last as in the 1st interval (Wells). In 45-sec. trials, the fatigue-index (loss of last 5 sec. divided by initial 5 sec.), according to Gilbert, is highest in young children (24 per cent at 8 years) and declines thence irregularly to 12.7 per cent at the age of 15. Tests by the author of fifty 8th-grade grammar-school boys reveal a fatigue-index (ratio of loss in 3d 10 sec. to 1st 10 sec. in 30 sec. tapping) of 13.7 per cent, m.v. 4.8 per cent, for the right, and 15 per cent, m. v. 4.6 per cent, for the left hand.

(*b*) According to Gilbert, the fatigue-index is higher for boys than for girls, but boys tap faster throughout each trial, so that their net efficiency is higher.[1]

(*c*) Kelly (10), who worked on a small number of children with a "fatigue-counter," found that "A"-grade pupils fatigued less than "C"-grade pupils; his index (the per cent of the last to the 1st 10 sec. in a 60-sec. trial) was for the former, with the finger 87.2 per cent, with the arm 88.0 per cent; for the latter, with the finger 77.0 per cent, with the arm 76.4 per cent. In the author's tests, no correlation could be discovered between fatigue-index and school standing.

(*d*) The effect of fatigue is progressively to 'level up' individual differences in speed. In other words, individual differences are more evident in initial than in terminal intervals (Bliss, Wells).

(*e*) Objective fatigue (slowing in rate) persists after practise, but the subjective feeling of fatigue may be eliminated thereby.

(*f*) The fatigue induced by 30-sec. tapping is apparently completely eradicated by a 3-min. rest-pause (Wells).

[1] It is more rea[illegible]e to interpret this higher index of boys as an expression of greate[illegible]nthusiasm than to follow Havelock Ellis in his inference th[illegible]ple of the "more continuous character of woman's activity" [illegible]Miss Thompson (17), from comparative tests of adults, [illegible]urpass women both in initial rapidity and in power [illegible] other tests show that high initial speed tends to be [illegible]tigue loss.

(*g*) The fatigue-index of right and left hands shows only slight correlation (Wells). The author's tests, however, show a correlation in the case of 50 boys of .33. In some persons the left hand is less susceptible to fatigue than is the right hand, though the reverse is the rule.

(*h*) The subjective experience of fatigue, as has been intimated, does not accord with the objective fatigue-loss.[1]

(10) *Practise* (*a*) The effect of practise is to p oduce a gradual improvement in speed, with, of course, occasional losses.

(*b*) The rise of the curve of efficiency is not, as in most activities, more rapid at the beginning than elsewhere.

(*c*) Maxima efficiency, when two experiments are performed daily, s reached, apparently, in about 20 days.

(*d*) Practise affects the left hand no more than the right; consequently the index of righthandedness is unaffected by repetition of the test.

(*e*) Practise particularly increases the rate in the later trials, *i.e.* it particularly affects warming-up, yet "the true practise gain is one mainly in the initial efficiency of performance, as distinguished from the warming-up gain, which shows itself chiefly in continued efficiency of performance" (Wells).

(*f*) An intermission of 10 to 14 days has no unfavorable effect upon practise gains, save that the feeling of fatigue may appear when work is resumed.

(11) *Diurnal rhythm.* Dresslar (7) found evidence of a diurnal rhythm with a minimum at 8 a.m. and a maximum at 4 p.m. Marsh (12 also found that afternoon records generally surpas ed those of the morning, though his figure do not accord very closely with those of Dresslar: Marsh also discovered that the later periods in the evening, which were not tested by Dresslar, furnished the most rapid rates of all.

[1] According to Wells: "The objective fatigue phenomena which we note in the test are in all probability either a fatigue phenomenon in the refractory phase or a lowered efficiency of coördination, especial[illegible] product of altered synaptic conditions; the sensations of fatigue, on th[illegible]and, may with equal assurance be ascribed to tissue changes with[illegible]es that take place as a result of their continued effort. In this t[illegible]he fatigue sensations are absolutely no indications of the a[illegible]litions, and any traceable correspondence between fat[illegible]atigue of performance must be regarded as almost wh[illegible] inhibition." (19, p. 473).

(12) *Dependence on 'general condition.'* When general well-being was ranked as good, medium, below medium, and poor, Wells was unable to discern any relation between these several conditions and tapping efficiency, while there was, in the case of susceptibility to fatigue, a tendency, if anything toward an inverse relation, *i.e.*, fatigue seemed to be greatest on 'good' days. Dresslar's observation that a vigorous walk decreases while mental work increases speed of tapping has been generally confirmed by other investigators.

(13) *Correlation with mental ability and social status.* (*a*) The correlation between tapping ability and mental ability is found to be generally positive by Smedley, Gilbert, Bolton, and Kirkpatrick, to be indifferent by Bagley (also by the author), while Binet and Vaschide report a positive correlation with 12-year old pupils and an inverse correlation with 16 to 20-year old pupils. While Gilbert found a very marked positive superiority of the 'bright' children in general, the relation did not appear at ages 16 and 17. Bolton found that "good children" (apparently meaning those drawn from the better social classes) were uniformly superior in tapping to children of the poorer class, both with the right and with the left hand. The fact that the divergence is greater at 9 than at 8 years, he attributes to a general arrest of development in the poorer-class children.

(*b*) Bolton also states that the "good" children showed a distinctly greater practise-improvement—a discovery which he terms "new and significant," and which he thinks is indicative of a fundamental difference in the ability o these two classes of children to take on new habits and profit quickly by experience.

(14) *Abnormal types.* Smith reports his inability to discern any characteristic differences between the speed of tapping in epileptics and in normal individuals, or between the speed of tapping and the rate of involuntary tremors in these cases. Wells' study of several cases of retardation in the insane (20), however, revealed alterations of the tapping activity, both in the form of a lowered average efficiency and also in the form of improvement in rate under conditions in which a normal individual would show either no change or a positive loss. These changes he terms 'reversal,' by which is meant intra- as distinguished from inter-trial warming-

up, and 'transference,' by which is meant a tendency for the index of righthandedness to be lower when the right-hand record is taken after the left-hand record.

(15) *Dependence on the type of movement.* The restriction of the tapping movement to specific joints, as has been attempted by some investigators, is difficult to accomplish in practise. However, it appears that the fastest rate is made when the movement is performed by the elbow joint, which is the one mainly concerned in the type of free movement here prescribed. Kelly, for instance, found that the speed of tapping was faster with the forearm than with the forefinger, in about the ratio 15 to 13. From this, in connection with other tests of dexterity, especially tests of minimal movement, he argues that children only gradually acquire dexterity and quickness of movement with the fingers, and that this passage "from fundamental to accessory," to use Burk's phrase, indicates the necessity of a general readjustment of the motor tasks required of children.

REFERENCES

(1) W. C. Bagley, On the correlation of mental and motor ability in school children, in A. J. P., 12: 1901, 193–205.

(2) A. Binet and J. Courtier, Recherches graphiques sur la musique, in A. P., 2: 1895 (1896), 201–222.

(3) A. Binet and N. Vaschide, (*a*) Épreuves de vitesse chez les jeunes garçons, in A. P., 4: 1897 (1898), 64–98. (*b*) Expériences de vitesse chez les jeunes gens, *ibid.*, 200–224.

(4) C. B. Bliss, Investigations in reaction-time and attention, in Yale S., 1: 1893, 1–55.

(5) T. L. Bolton, The relation of motor power to intelligence, in A. J. P., 14: 1903, 615–631.

(6) W. L. Bryan, On the development of voluntary motor ability, in A. J. P., 5: 1892, 123-204.

(7) F. B. Dresslar, Some influences affecting the rate of voluntary motion, in A. J. P., 4: 1892, 514–527.

(8) J. A. Gilbert, Researches on the mental and physical development of school children, in Yale S., 2: 1894, 40–100.

(9) J. A. Gilbert, Researches upon school children and college students, in Iowa S., 1: 1897, 1–39.

(10) R. L. Kelly, Psychophysical tests of normal and abnormal children; a comparative study, in P. R., 10: 1903, 345–372.

(11) E. A. Kirkpatrick, Individual tests of school children, in P. R., 7: 1900, 274–280.

(12) H. D. Marsh, The diurnal course of efficiency, Columbia Univ. diss.' N. Y., 1906. Pp. 99.

(13) J. M. Moore, Studies of fatigue, in Yale S., 3: 1895, 68–95.

(14) O. Raif, Ueber Fingerfertigkeit beim Clavierspiel, in Z. P., 24: 1900, 352.

(15) F. Smedley, Rept. dept. child-study and pedagogic investigation, No. 3, 1900–1901 (Chicago Public Schools). Also in U. S., 1902, i, 1115–1138.

(16) W. G. Smith, A comparison of some mental and physical tests in their application to epileptics and to normal subjects, in B. J. P., 1: 1905, 240–260.

(17) Helen B. Thompson, The mental traits of sex, Chicago, 1903. Pp. 188.

(18) F. L. Wells, A neglected measure of fatigue, in A. J. P., 19: 1908, 345–358.

(19) F. L. Wells, Normal performance in the tapping test before and during practise, with special reference to fatigue phenomena, in A. J. P., 19: 1908, 437–483.

(20) F. L. Wells, Studies in retardation as given in the fatigue phenomena of the tapping test, in A. J. P., 20: 1909, 38–59.

(21) F. L. Wells, Sex differences in the tapping test: an interpretation, in A. J. P., 20: 1909, 353–363.

TEST 11

Accuracy or precision of movement: Aiming.—The purposes for which tests of accuracy of movement have been employed are practically the same as those cited for the tapping test, viz: to obtain an index of general voluntary motor ability with which to compare different children to compare the right with the left hand, to determine the development of motor control with age, its differentiation with sex, and to test its correlation with mental ability. These tests have been only rarely used for determining the presence of fatigue, though they have been proposed as means for the diagnosis of incipient ataxia.

Tests of accuracy vary greatly in form: in fact, they virtually shade by degrees from those which prescribe a rapid accurate movement similar to the tapping movement of Test 10, *e.g.*, the Columbia test described by Wissler (5), to those which prescribe a slow steady movement more akin to a test for steadiness (No. 13).

Two types of precision test have been selected for consideration, the aiming test and the line drawing or 'tracing' test (No. 12).

The common feature of all forms of aiming test is the measurement of the extent of error made by an individual when he tries in a series of discrete voluntary movements of hand or arm to hit some form of mark or target. According to the particular form of movement employed, the test has been known as the 'probing test,' the 'target test,' the 'thrusting test,' etc. These movements

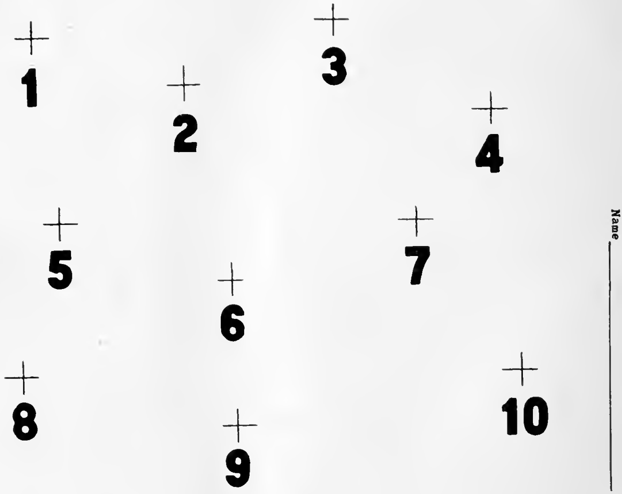


FIG. 23. TARGET BLANK.

The numbers are added to show the order in which the crosses are to be struck. Cut ½ size.

have ranged, to speak more specifically, from a simple vertical probing movement of 6 mm. extent (Bryan, 2) to whole-arm aiming with a pencil at a paper target at arm's length (Thompson, 3; Whipple, 4), or making lunging thrusts with a wand, somewhat after the fashion of a fencer, or even throwing ordinary marbles at a bull's eye target 2 m. distant (Bagley, 1).

The results of any such precision test will obviously be condi

tioned by the position of the target with respect to S, by the extent and rate of the aiming movement, and likewise though the fact seems not always to have been recognized, by the individually variable improvement in accuracy which will appear if a series of 'shots' are taken at the same target. Hence, to be satisfactory, an aiming test should prescribe and standardize all these conditions: it should also admit of an exact evaluation of each aiming movement. The form of test here described was devised by the author severa years ago to meet these conditions and has proved satisfactory in use. Though the error of a single stroke is large (as

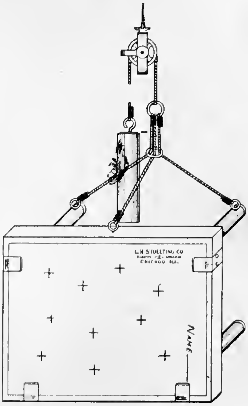


FIG 24. ADJUSTABLE BASE-BOARD FOR TARGET-TEST.

is certain to be the case in any form of aiming test), the average of the 30 thrusts made by the same S is very constant. The use of ten marks in place of a single mark, or bull's eye, removes to a large extent the improvement error just mentioned.

Apparatus.—Prepared blanks containing ten crosses irregularly arranged (Fig. 23). A base board upon which the blanks are fastened, arranged to be secured upon the wall and adjusted to varying heights (Fig. 24). Metronome (Fig. 14). Pencil with tough and moderately hard lead. M llimeter scale.

PRELIMINARIES.—Fasten the board upon the wall and arrange the counterweight properly, so that the board will remain in any position from one to two meters from the floor and will not be displaced when struck by the pencil.

Set the metronome at 69. Fasten the target-sheet upon the board, with the name-date corner n the lower right-hand corner of the board Place a demonstration target on the wall conveniently near.

METHOD.—Make clear to *S* the following directions. (1) He is to stand with his ight shoulder (for the right-hand test) squarely in front of the target, at such a distance that his pencil just strikes the target when his arm is fully extended.[1] (2) He is to strike in time with the beat of the metronome. (3) Each stroke is to be a full, smooth stroke, not jerky or too short, and the pencil must, therefore, be brought back, each time, until it touches the shoulder. (4) He is to start at the first cross and make successive strokes, one at each cross in the series until the tenth is reached (see Fig. 23). This process is twice repeated, but in the second round, further to avoid practise, the order is from ten to one. *S* thus makes three shots at each mark, or thirty in all.

Before conducting the test proper, let *S* try the experiment upon the demonstration target. It will save time if *E* also illustrates the process at this time. *E* should count out the strokes of the metronome: 'one, two, three," etc., to assist *S* in getting the proper rate.

Place a fresh sheet upon the target-board and test the left hand.

TREATMENT OF RESULTS.--Measure the error of each thrust by the application of the millimeter scale. A pair of dividers may be helpful in this process. Average the thirty errors and compute the mean variation or standard deviation. Any 'shots' that have struck the lines of the crosses and are difficult to detect may be easily located by reversing the sheet.

RESULTS.—(1) On the basis of similar tests, other investigators have shown a gradual increase in accuracy with *age*, particularly during the years 5 to 8.

[1] If the subjects are of approximately the same size, this distance may be marked upon the floor by a chalk line.

(2) *Sex* differences are slight, but on the whole boys are more accurate than girls, and men than women.

(3) The author, in using the test as described, has found an error of 4 to 6 mm. in university students, while in a group of fifty 8th-grade boys the following results were obtained:

Right hand, average 5.12, lowest 3.75, highest 8.34.

Left hand, average 6.39, lowest 4.15, highest 9.27.

(4) For the 50 boys just mentioned, the correlation between right and left-hand efficiency was 0.54.

REFERENCES

(1) W. C. Bagley, On the correlation of mental and motor ability in school children, in A. J. P., 12: 1901, 193–205.

(2) W. L. Bryan, On the development of voluntary motor ability, in A. J. P., 5: 1892, 123-204.

(3) Helen B. Thompson, The mental traits of sex, Chicago, 1903. Pp. 188.

(4) G. M. Whipple, The influence of forced respiration on psychical and physical activity, in A. J. P., 9: 1898, 560–571.

(5) C. Wissler, The correlation of mental and physical tests, in P. R. M. S., 3: No. 6, 1901. Pp. 62.

TEST 12

Accuracy, precision, or steadiness of movement: Tracing.—The purposes for which tracing has been used are identical with those outlined for the preceding test, but the present test differs from the former in that the movement is continuous, analogous to that made in drawing a line—the so-called 'writing movement' of Bryan (3) or 'tracing test' of Bagley (1). Since steadiness of movement is quite as much in demand as accuracy (of the sort required in Test 11), this test is often classed as a steadiness test, rather than as an accuracy test, but it differs from steadiness tests proper in that it measures control of a voluntary movement, whereas the latter measure the extent of involuntary movement which takes place when the hand or arm is held at rest (Test 13).

The technique most commonly adopted for the tracing or line-drawing test consists in passing a metallic needle or stylus along a narrow slit between metallic strips and noting by telegraph sounder, bell, electric counter, or graphic record, the number of contacts

made in passing along a given portion of the slit. This slit may be straight and bounded by parallel sides (Bolton, 2) or by slightly converging strips (Bryan, 3; Thompson, 4), or the slit may in portions be curved, as in the scrolls used by Bagley. Some tests have been made at Columbia University with an irregular printed pattern, which is to be traced by the subject with a lead pencil. In any of these tests, the movement may be arranged to involve primarily either the finer muscles of the hand and fingers or the larger muscles concerned in the whole-arm movement.

The test here described follows the method used by Bryan and by Thompson.

Apparatus.—Tracing-board (Fig. 25). Metallic stylus with flexible connecting wire. Telegraph sounder (Fig. 26). Two dry or other open-circuit cells. No. 18 annunciator wire.

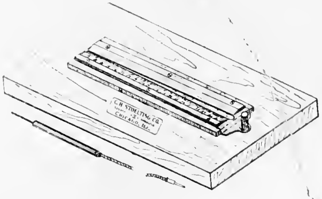


FIG. 25. TRACING-BOARD.

Preliminaries.—Wire the battery in series with the tracing-board, the sounder, and the stylus.

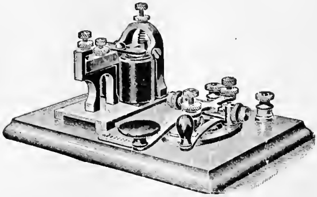


FIG. 26. TELEGRAPH SOUNDER AND KEY.

The sounder may be purchased separately, but the key will be found useful for other experimental work, if only for opening and closing the circuit at will. When provided with a special pointer, this sounder is used for graphic records, as in Test 13.

Method.—(1) Seat S comfortably with the tracing-board squarely before him, and the apex of the angle pointing toward him, so that the movement is directly toward the body in the median plane. Let S hold the stylus as he chooses, place the tip at the opening of the strips, and then attempt to draw a line on the glass between the strips of metal without touching either one. The movement should be *continuous* from start to finish, made entirely free-arm (not a finger or wrist movement and without supporting the hand or arm in any way.) The rate of movement must be illustrated as accurately as possible by E, and should be such that the full length of the strips is traversed in 9 sec. Allow S two or three preliminary trials, and endeavor to secure approximately this rate before starting the records. As soon as a click of the sounder indicates a contact, S is to stop and begin again with the left hand. Repeat until S has made five trials with each hand, alternately. E records in each case the point on the scale at which contact is made.

(2) Turn the apparatus around 180°, and in the same manner test movement away from the body.

(3) Test movement from left to right and (4) from right to left with either hand, by placing the test-board so that the strips lie parallel with the edge of the table nearest S.

Variations of Method.—(1) If it is desired not to compare movements in different directions, but merely to compare the right and left-hand efficiency of different S's, the test may with advantage be shortened by adopting the method used by Bryan in his tests of school children, viz: set the test-board so that the strips make an angle of approximately 45° with the edge of the table, *i.e.*, with the right-hand end of the strips turned 45° away from the body for the right-hand test, and with the left-hand end of the strips turned away to the same amount for the left-hand test. Make five tests with movement inward and five with movement outward with each hand.

The conditions may be still further varied (2) by requiring S to stand and to hold the stylus at arm's length, (3) by allowing S to sit and to support his arm at the elbow, or (4) to support his hand on the base-board while executing a forearm or whole-arm movement. The last was the method followed by Miss Thompson.

Treatment of Results.—The simplest treatment of the data is to secure an index of precision by averaging the distances at which the several points of contact are made. For a more complex method of computing the measure of precision, the reader is referred to Bryan, pp. 180 ff.

Results.—(1) There is greater *variation* in the outcome of precision or steadiness of movement tests than in that of rate of movement tests.

(2) There is undoubtedly a more or less constant improvement in precision with *age*, but sufficient data are not yet at hand to determine the yearly increments clearly. There is certainly, however, a decided gain during the years 6 to 8, while Bolton also noted improvement from the 8th to the 9th year of age.

(3) *Dextrality.* In general, the right hand is, of course, distinctly superior to the left, but the amount of this superiority varies remarkably with age, and is, according to Bryan, less evident at 15 and 16 than at 6, 9, or 12 years of age.

(4) *Sex.* With either hand, boys are probably slightly superior to girls, for, while Bolton reports the superiority of the girls in his groups of children, Bryan's examination of some 700 children revealed the following relations, when the results for both hands were computed together: boys were superior to girls in 51.5 per cent of the trials, girls superior to boys in 35.3 per cent, and the sexes equal in 13.4 per cent. Moreover, in a similar test, Thompson found men superior to women.

(5) Both Bolton and Thompson found *movements inward* or toward the body uniformly steadier than movements outward or away from the body.

(6) *Correlations.* Bagley found "a decidedly inverse relation between mental ability, as indicated by class standings, and motor ability, as indicated by the tracing test." But Bolton, who used a different form of test, and apparently correlated rather with the social status than with the class standing, reports that 'good' are steadier than 'poor' children.

REFERENCES

(1) W. C. Bagley, On the correlation of mental and motor ability in school children, in A. J. P., 12: 1901, 193–205.

(2) T. L. Bolton, The relation of motor power to intelligence, in A. J. P., 14: 1903, 615–631.

(3) W. L. Bryan, On the development of voluntary motor ability, in A. J. P., 5: 1892, 123-204.

(4) Helen B. Thompson, The mental traits of sex, Chicago, 1903. Pp. 188.

TEST 13

Steadiness of motor control: Involuntary movement.—The general idea in this type of test is to measure the amount of involuntary movement which appears when the finger, hand, arm, or body as a whole is held as nearly motionless as possible. Like the preceding tests, this has been frequently employed as a means of obtaining an index of motor ability (Bagley, 1; Bolton, 2; Hancock, 6), but it has also been employed for numerous other purposes, *e.g.*, for examining the motor tendencies accompanying ideational activity (Jastrow, 7, 8; Tucker, 15), for examining the bodily expression of affective states (Titchener, 14), for examining the nature of constant tendencies toward automatic movements, or the possibility of developing such movements by training (Thompson, 13; Solomons and Stein, 10; Stein, 12), and for detecting the presence of incipient or recent chorea (Crichton-Browne, 4).

This variety of purpose illustrates very forcibly the difficulty of so conducting the test as to examine any one phase alone—a difficulty which is aggravated by the fact that the tracings of involuntary movement are often affected, not only by the factors implied above, but also to a considerable extent by the direction of the attention, by the relative position of the body and the instrument, and by physiological processes, especially respiration. It is, therefore, not surprising that many writers, *e.g.*, Bagley and Bolton, have incorporated records of involuntary movement only with qualifications and without placing much insistence on their worth.

The instruments most commonly employed are the ataxiagraph (described by Crichton-Browne, used by him and by Hancock and Bolton for measuring the swaying of the body as a whole), the tremograph (described by Bullard and Brackett, 3, and also used by Hancock) for testing the arm or finger, the automatograph of Jastrow or that of Stein, further elaborated by Titchener, for measuring involuntary movements of the arm, the digitalgraph

devised by Delabarre (5), described, together with others of the instruments just named, by MacDonald (9), and used for recording the tremor of the finger, and several instruments, as yet unnamed, for testing either the arm or hand, according to the conditions of their use.[1]

APPARATUS.—Brass plate, set at an angle of 45°, and pierced with a series of holes whose diameters are 32, 20, 16, 13, 11, 10, 9, 8, and 7 sixty-fourths of an inch, respectively (Fig. 27). Metallic needle of special design, with flexible connecting wire. Telegraph

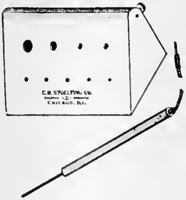


FIG. 27. STEADINESS TESTER.

sounder (Fig. 26), with writing lever attached to the armature, and with sending key (or separate short-circuiting key). Kymograph (Fig. 15), with accessories. Stop-watch (Fig. 13). Four dry or other open-circuit cells. Insulated connecting wire. [A low table, about 65 cm. high and an adjustable chair of the typewriter chair pattern are desirable.]

PRELIMINARIES.—Wire the needle, brass plate, sounder and key in series with the battery in such a way that contact between the needle and the plate will actuate the sounder. Place smoked paper on the kymograph drum, and adjust to approximately one

[1] For some purposes it is probably quite as well to test the subject without the use of apparatus. Thus, according to Sturgis, the following constitutes an infallible test for chorea: "Bid the child hold up both hands open, with extended arms, the palms toward you. If that is done steadily, both hands upright and both alike, no finger or hand quivering, no falling back of either hand, nothing to choose between the positions of the two, then the child has not, nor is it near (either before or after) St. Vitus' dance. You may confirm this test by another. Let the child place the open hands upon yours, palm to palm. Look then at the backs of the child's hands, observe whether fingers or thumbs (especially the latter) repose without tremor and without restraint."

revolution in 40 sec. Adjust the position of the sounder so that the writing lever gives a satisfactory tracing on the drum.[1]

Method.—Seat *S* before the table in a comfortable position. Place the brass plate flush with the front edge of the table, in front of *S*'s right shoulder for the right-hand test, and in front of his left shoulder for the left-hand test. Instruct *S* to hold the needle in such a way that his finger tips are in contact with the expanded flange of the holder, and at command, to hold the tip of the needle within the largest hole, and to maintain this position, so far as possible, without touching the brass plate during the 15 sec. allowed for the trial. *S*'s hand and arm must be entirely free from all support or contact with his body or other object, and his forearm should form an angle of approximately 100° with his upper arm. The needle should be inserted about 6 mm. into the hole.

Show *S* that the click of the sounder will serve as warning for him that the needle is making contact with the plate.

In conducting the test, allow *S* about 3 sec. for taking the position (since a certain amount of movement will appear when the needle is first inserted that will afterward be checked by *S*'s control), then close the short-circuiting key and at the same time start the stop-watch. At the expiration of 15 sec., open the circuit, stop the watch and the kymograph, and at once rearrange the instruments for the left-hand test.

Allow 30 sec. rest, and then test the right hand and the left hand with the next smaller hole, and so on, until a hole is reached so small that *S* has reached the limit of his capacity, and is clearly unable to keep the needle free from contact with the plate.

In most cases a few unrecorded preliminary trials will show approximately the degree of *S*'s control, and the tests with the larger holes may be omitted with a consequent saving in time.

Variations of Method.—The test may be modified by altering the relative position of the plate and *S*'s body as follows: (*a*) by requiring *S* to stand and to hold the needle extended at arm's length, (*b*) by allowing *S* to rest his elbow upon the table with the forearm free, or (*c*) by supporting his forearm and wrist and testing the steadiness of the hand and fingers, and (*d*) by extending the time to 30 or 60 sec. The results will naturally differ character-

[1] For an account of the manipulation of the kymograph, consult Test 10.

istically from those obtained under the conditions prescribed as standard.

TREATMENT OF RESULTS.—By reference to the kymograph tracings, count the number of contacts for each hole. For comparative purposes, *E* may take the total number of contacts made in a given series of holes or the number made in that hole which most satisfactorily tests the steadiness of the subjects under investigation.

RESULTS.[1]—(1) In all tests of involuntary movement, it is clearly seen that *age* is an important factor. Hancock concludes that adults have approximately 5.8 times as much control over their fingers as do children aged 5 to 7 years.

(2) Distinct *sex* differences have not been established.

NOTES.—In rare cases, the use of the graphic method may be dispensed with, especially if *E* proves by practise very skillful in the correct counting of the rapid and irregular strokes made by the sounder when the test approaches the limit of *S*'s capacity, but this simplification is not recommended, because it is exceedingly difficult to make the count under these conditions (cf. Bolton), and moreover, there may appear very short rapid contacts that will actuate the sounder sufficiently to produce a noticeable indication on the tracing, but not sufficiently to produce a noticeable click. Again, *S* will occasionally make contacts of longer duration: with the graphic record, it is possible for *E* to measure the duration of these contacts, and, if desired, to base *S*'s record upon the proportion of the time during the 15 sec. that is occupied in contact, rather than upon the number of contacts alone.

The electric counter is not recommended as a substitute for the graphic method, because, as mentioned in Test 10, it will not operate reliably when the contact is very brief, even although a large number of cells are used in the battery. To avoid the possibility of a similar error in the use of the sounder, the excursion of the armature should be rather short, *i e.*, the armature should be adjusted as near the fields as is possible if it is to give a clean stroke.

If involuntary tremor is to be studied with special care, *e.g.*, if *E* wishes to make an extended study of an individual case, a

[21] A summary of the results obtained with the automatograph is given by Jastrow (8, pp. 307-336).

more sensitive instrument should be employed. For such work, the tridimensional analyzer of Sommer (11), also described by Titchener (14), is recommended.

REFERENCES

(1) W. C. Bagley, On the correlation of mental and motor ability in school children, in A. J. P., 12: 1901, 193–205.

(2) T. L. Bolton, The relation of motor power to intelligence, in A. J. P., 14: 1903, 615–631.

(3). Bullard and Brackett, Boston Med. and Surg. Jour., 2: 1888, 600.

(4) J. Crichton-Browne, The nervous system and education, being part 4 in The book of health, edited by M. Morris, London, 1883 (now out of print).

(5) E. B. Delabarre, Ueber Bewegungsempfindungen, 1891. See also A. P., 1: 1894 (1895), 532.

(6) J. A. Hancock, A preliminary study of motor ability, in Pd.S., 3: 1894, 9–29.

(7) J. Jastrow, (*a*) A study of involuntary movements, in A. J. P., 4: 1892, 398–407. (*b*) A further study of involuntary movement, *ibid.*, 5: 1892, 223–231.

(8) J. Jastrow, Fact and fable in psychology, Boston, 1900. Pp. 370.

(9) A. MacDonald, Experimental study of school children, etc., chs. 21 and 25 in U. S., 1897–8 (1899).

(10) L. Solomons and G. Stein, Normal motor automatisms, in P. R., 3: 1896, 492–512.

(11) R. Sommer, Lehrbuch der psychopathologischen Untersuchungs-methoden, Berlin, 1899. Pp. 399.

(12) Gertrude Stein, Cultivated motor automatisms, in P. R., 5: 1898, 295–306.

(13) Helen Thompson, The mental traits of sex, Chicago, 1903. Pp. 188.

(14) E. B. Titchener, Experimental psychology, 2 vols., N. Y., 1901 and 1905.

(15) M. A. Tucker, Comparative observations on the involuntary movements of adults and children, in A. J. P., 8: 1897, 394–404.

CHAPTER VI

Tests of Sensory Capacity

Psychophysical tests of sensory capacity are divisible into tests of liminal sensitivity (sensitivity proper) and tests of discriminative or differential sensitivity (sensible discrimination.)[1] In the former, we measure the bare capacity of experiencing sensations, the minimally perceptible stimulus or *stimulus limen, e.g.*, the lightest pressure that can be felt, the least intensity of tone that can be heard, etc; in the latter, we experience different sensations and report upon their difference; we seek, in other words, to determine the minimal objective difference of stimulation that can just be mentally cognized as different, to determine the *difference limen, e.g.*, the smallest change of vibration-rate that will suffice to yield two perceptibly different tones, or the smallest difference in weight that can just be recognized as a difference.

These two measurements of sensory capacity, liminal and discriminative sensitivity, can be applied to any modality, *i.e.*, to any sense-department, and to any attribute of sensation, *i.e.*, to quality, intensity, extent, and duration. We may measure, for instance, in the case of the ear, liminal sensitivity in terms of intensity, discriminative sensitivity in terms of intensity, in terms of quality, in terms of duration, etc. We may, furthermore, determine the total number of auditory sensations that can be experienced, *i.e.*, modal sensitivity.

In the case of the two sense-departments that possess the attribute of extent, visual and cutaneous sensations, we may also measure the capacity to discriminate difference in the localization of two stimuli, or the limen for spatial discrimination. While, strictly speaking, this may be regarded as a more complex process than simple sensory discrimination, it may, for our purposes, be included

[1] For a discussion of the terminology, methods, purposes, and results of psychophysical methods, consult Titchener, *Experimental Psychology*, especially, Vol. 2, part 2.

as a sensory test. Indeed, with both the eye and the skin, this determination, for practical purposes of exploration of the sense-organ or of the measurement of its functional capacity, has superseded the determination of liminal sensitivity. Thus, in visual sensation, visual acuity refers, not to the liminal sensitivity of the retina for stimulation, but, in principle, to the capacity to distinguish the separation of two points. Similarly, in cutaneous sensation, for a practical test of functional capacity, the determination of the 'limen of duality,' as it may be termed, by means of the esthesiometer, has been more often employed than the simple determination of liminal sensitivity to pressure.

Now, the quantitative determination of sensitivity in the psychological laboratory has given rise to a most elaborate and refined methodology, and has, in fact, been the chief problem of the science of psychophysics. It is not the purpose here to discuss or duplicate these exact methods, but merely to indicate the manner in which, for comparative purposes, one may secure an index of functional efficiency by empirical methods. It must be clearly understood that the determination of an exact stimulus or difference limen in the psychological laboratory, with minute introspective analysis of the factors that condition the process and with elaborate methodological procedure, is quite a different process from this simple determination of functional capacity for comparative purposes. If, for convenience, the technical terms of psychophysics are here employed, they are employed with this qualification in mind.

To make this point clearer, the procedure for the determination of discriminative capacity as herein recommended is not identical with any of the established psychophysical methods. *E* allows *S* at first to try various stimulus differences ranging from large to small, until *S* has acquired general familiarity with the test, and *E* has obtained a general notion of the *S*'s capacity. *E* may next test *S* more formally by applying a series of stimulus differences ranging from clear subjective difference to subjective equality. He then selects a difference which seems likely to be just cognizable by the subject and applies this difference ten times, with proper reversals for time or space errors. If eight right judgments are given, he then corroborates the result by trying similarly a slightly smaller,

and finally a slightly larger difference, to see if *S* gives in the former case fewer, and in the latter case more correct judgments. *S* knows that a difference exists, but is ignorant of its spatial or temporal position.[1] We thus obtain an index of capacity, but do not determine the mean difference limen, nor even the lower limen, in the psychophysical sense.

Sensory tests of this empirical sort have been employed, partly in connection with the psychology of individual and sex differences, partly in the objective study of general intelligence, partly in the exploration of sense organs for the determination of their working condition, *i.e.*, for hygienic and diagnostic purposes. In all of these fields the emphasis is upon the examination of simple functional capacity, without particular reference to introspective examination or analysis of the accompanying consciousness.

The use of sensory tests in correlation work is particularly interesting. In general, some writers are convinced that keen discrimination is a prerequisite to keen intelligence, while others are equally convinced that intelligence is essentially conditioned by 'higher' processes, and only remotely by sensory capacity,—barring, of course, such diminution of capacity as to interfere seriously with the experiencing of sensations, as in partial deafness or partial loss of vision.

While it is scarcely the place here to discuss the evolutionary significance of discriminative sensitivity, it may be pointed out that the normal capacity is many times in excess of the actual demands of life, and that it is consequently difficult to understand why nature has been so prolific and generous; to understand, in other words, what is the sanction for the seemingly hypertrophied discriminative capacity of the human sense organs. The usual 'teleological explanations' of our sensory life fail to account for this discrepancy. Again, the very fact of the existence of this surplus capacity seems to negative at the outset the notion that sensory capacity can be a conditioning factor in intelligence,—with the qualification already noted.

The tests which follow are selected from a large number of theoretically possible tests, because of their prominence in such experimental studies as have been mentioned. Their classification is simply by sense-departments. Tests for the exploration of the organ, measurement of its defects, determination of acuity, liminal

[1] Procedure with 'half-knowledge' in the sense used by Kämpfe, in Phil. Stud., 8: 1893, 543, and Wundt, Grundzüge d. physiol. psych., 5th ed., i., 1902, 492. See also Titchener, ii, Pt. 2, 127.

sensitivity, discriminative sensitivity, in so far as they are described, are given successively for each of the main sense departments.

TEST 14

Visual acuity.—The functional capacity of the eye is examined primarily, of course, for practical purposes in connection with hygienic investigation. Occasionally, it becomes desirable to determine the presence or absence of visual defect in connection with the administration of some mental test, *e.g.*, the cancellation test.

Visual acuity has been studied in its relation to school standing, general intelligence, occupation, habitat, race, sex, as well as to bodily disturbances, such as headache, chorea, indigestion, or other optical defects, such as strabismus, total color-blindness, etc.

Optical inefficiency, aside from color-blindness, may be due to *amblyopia* (dimness of sight not due to refractive errors or demonstrable lesion), or to *asthenopia* (general impairment of retinal efficiency due to anæmia, over-use, etc., and often yielding to proper medical treatment), but is more commonly some form of *ametropia* (defect in shape of the eye-ball, lens, or cornea, with resultant defect in refraction and in the formation of the retinal image).

Ametropia may exist as presbyopia, myopia, hyperopia, or astigmatism.

Presbyopia is the long-sightedness of old age, due to the lessened elasticity of the lens.

Myopia, or short-sightedness, is commonly produced by too long an eye-ball, the effect of which is to allow rays of light in distant vision to focus in front of the retina and hence to produce a blurred image when they finally impinge upon the retina. The myopic eye is thus unable by any effort clearly to see objects situated at distances of 2 m. or more away, while its 'near-point,' *i. e.*, the nearest point at which clear vision is possible, is brought correspondingly closer, so that objects may be seen clearly when 5 or 6 cm. distant,—something which would be impossible for the normal (emmetropic) eye. Myopia is rarely congenital, but is an acquired defect, and characteristically a disease of civilization and culture. Pure myopia, as a rule, causes no eye-strain, but it is nevertheless a serious condition, because of its tendency to increase in degree, and because of the appearance in many cases of concomitant pathological disturbances of the retina, which, in extreme cases, result in actual blindness. In practise, moreover, myopia is

rarely found pure, but complicated with astigmatism and other defects. It may be counteracted, and its progress checked, but not cured, by the use of properly fitted concave lenses, supplemented by the exercise of caution in the use of the eyes.

Hyperopia, hypermetropia, or long-sightedness, more exactly over-sightedness, is commonly produced by too short an eye-ball, the effect of which is to intercept rays of light too soon, *i.e.*, before they are brought to a normal focus. The hyperopic eye must consequently exert an effort of accommodation in order clearly to see objects at a distance, while for near work this effort must be excessive. The result is that the hyperopic eye is under constant and abnormal strain from the incessant demands upon its ciliary muscle, and that, in consequence, numerous secondary symptoms or resultant effects appear, some of them obvious, others unexpected, many of them serious. Local symptoms appear in inflammation, redness, or soreness of the eyes, lids, or conjunctiva, and in twitching and pain within the eye-ball. Aside from these local disturbances, perhaps the most constant symptom of hyperopia is frontal or occipital headache. Characteristic also is the holding of the work at some distance from the eyes,[1] a peering or frowning expression, and dislike of near work. Eye-strain, whether hyperopic or astigmatic, may also occasion more serious physiological disturbances, such as chorea, vomiting, nervous dyspepsia, etc.[2] Since the hyperopic eye can see clearly at a distance and can read (as its possessor often boasts) with the book held at some distance, the defect is often unsuspected, because the secondary symptoms are not correctly interpreted. On this account, too, it becomes necessary to take special steps to detect its presence, and many of the simple distance tests that have been applied wholesale upon school children utterly fail to diagnose it. The oculist commonly makes use of homatropin or some other cycloplegic to paralyze temporarily the ciliary muscle and thus prevent accommodation. Hyperopia may, however, be detected, though less accurately, by the use of suitable test-lenses, as described below. The defect is counteracted by the use of properly fitted convex lenses.

Astigmatism is produced by an uneven radius of curvature, usually of the cornea: this surface, which should normally be approximately spherical in form, is, in astigmatism, more strongly curved in one axis or meridian than in another, so that the cornea is ellipsoidal in form, *e. g.*, like the bowl of a spoon, or the side, rather than the end of an egg. Thus the eye is double-focussed, and it is impossible by any effort to focus an image clearly in both meridians simultaneously. In measuring astigmatism it is evident that one must assign both the degree of refractive error and the axis in which the

---

[1] In high grades of hyperopia, distinct images can not be secured even by this process, so the child may abandon the attempt to secure clearness and seek merely to increase the size of the image by holding his book near his eyes. He may thus be falsely rated as near-sighted by the casual observer.

[2] The injurious effects of eye-strain have found a special expositor in Dr. G. M. Gould (6, 7).

error lies, and that in correcting it, a cylindrical lens of the proper curvature must be placed before the eye at exactly the proper axis to counteract the indicated deficiency. This lens only counteracts the defect, and does not cure it. Astigmatism may be in part congenital, in part a phenomenon of growth (often attributable to the pressure upon the eye-ball of the eye-lids and contracted brows, with the result that the maximal refractive index lies at or near the vertical meridian). When present in large amount it becomes a serious obstacle to vision; when present in small amounts, as is apt to be the case in many eyes, it is the occasion of the same phenomena of eye-strain that have been mentioned as accessory to hyperopia; astigmatic headache is particularly symptomatic,—indeed, 60 per cent of all headaches are said to be traceable to this source.

It must be understood that these three defects may, and commonly do, appear in combination, particularly astigmatism with hyperopia or myopia, and that the defects may be, and commonly are, unlike in the two eyes of the same individual. Partly for this reason, the proper fitting of glasses is an art, and, like any art, requires great skill, complete familiarity with the conditions, and long practical experience. The tests which are here described make no pretence to exactitude, but are designed to determine, in so far as is possible by simple methods, the existence of defects that *should invariably be referred to a specialist for further examination and treatment.*

For the examination of refraction the chief appliances are (1) the ophthalmometer, for the exact measurement of the degree and axis of astigmatism, (2) the ophthalmoscope, for the examination of the rètina, (2) the retinoscope and the skiascope, for the objective determination of refractive errors, (4) test-types and trial lenses, for actual visual tests under varying conditions. While retinoscopy is a method of great value, especially in testing young children, the test-type is, in general, the court of final appeal and constitutes the most widely used and perhaps the most valuable single means for testing visual acuity. The most varied kinds of test-type have been devised by oculists. Perhaps oldest and best known are Snellen's "Optotypi," which form the basis of the tests ordinarily used. Interesting variations are seen in Dennett's Monoyer type, Landolt's C-test, and Cohn's E-test.

The simple test which is described just below is recommended by the American Ophthalmological Society, and is designed to be used in connection with Dennett's type, with the employment of but two test lenses. The second test supplements the first by detecting astigmatism. Tests 15, 16, and 17 may be added.

A. TEST FOR AMETROPIA

Apparatus.—Dennett's Monoyer test-type (Fig. 28).[1] Trial frame (Fig. 29). Two − .75 D. and two + .75 D. spherical lenses, and one blank disc.

FIG. 28. DENNETT'S MONOYER TEST-TYPE.

About $\frac{1}{7}$ natural size. The small figures that indicate the normal distance for each line are not shown.

[1] Other test-types may, of course, be employed. Landolt's or Snellen E-test (designed for illiterates) are recommended, if Dennett's is not used. Landolt's type is placed at 5 m., Snellen's at 6 m. Visual acuity is indicated in the latter by a fraction in which the numerator is 20 (feet) and the denominator is the number (in feet) indicated for the smallest line that can be read at the standard distance of 20 feet.

PRELIMINARIES.—Place the test type on the wall or stand, on a level with *S*'s eyes, in a strong even illumination, though not in actual sun-light.[1] Seat *S* comfortably at a distance of 6 m. from the chart.

Note any indication of soreness or inflammation of the eyes, lids, or conjunctiva. Ascertain if *S* has ever suffered from such inflammations, from habitual headaches, or watery eyes; whether his eyes become painful, sore, or strained in doing close work; whether he

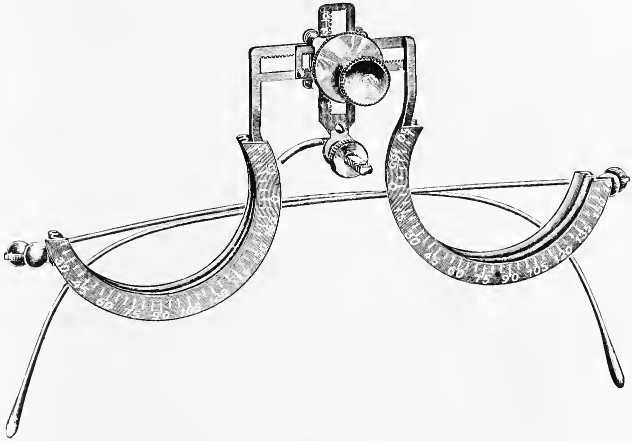


FIG. 29. TRIAL FRAME.

Holds two pairs of lenses, and has rack and pinion adjustment for pupillary distance, and for vertical, and back and forward movement of the nose-piece.

has previously been examined, and if so, with what result; whether he has ever worn glasses; if they have been worn and discarded

[1] If conditions render daylight illumination unreliable or unsatisfactory, *E* must arrange artificial illumination, carefully shaded from *S*. Excellent devices for this purpose may be purchased from dealers in optical supplies, *e.g.*, No. 4274, catalog of E. B. Meyrowitz, New York City, price $10. Whatever the source of illumination, the light must not shine in *S*'s eyes.

An experienced *E* may compensate for inadequate illumination by placing *S* at a distance less than 5 m.,—the amount of the correction being determined by *E*'s own acuity under the prevailing conditions.

ascertain when and why. If S wears glasses, record that fact; test his vision both with and without his glasses, unless time forbids, in which case test with them.

Adjust the pupillary distance and the nose-piece of the trial frame so that the lenses will be centered before S's eyes.

METHOD.—(1) Place the solid disc in the frame before S's left eye. Instruct him always to keep both eyes open. Ask him to read the top or smallest line. If this line can be read entire, or at least 11 of the 13 letters can be read, record the vision of the right eye as 1. (indicated by the figure at the right of the line), thus, V. R. E. = 1. If he sees nothing clearly above, say the 4th line, but can read that correctly, then record, as indicated on the chart, V. R. E. = .7. If S cannot read any of the lines, record V. R. E. = − .7, which is practically blindness.

(2) *Whatever* the result of this first trial, always next place the + .75 D. lens in the trial frame. This will blur the vision if the eye is emmetropic, so that, if before, V. R. E. = 1., and if vision is now blurred, record V. R. E. = 1. Em.

If S can, with the plus lens, read the same line as before, or a smaller line than before, then the eye is hyperopic. Thus, if previously the 4th line was read and now the 2d, the record will be, V. R. E. = .7 + Hy. = .9, or, if no improvement appeared, V. R. E. = .7 + Hy. = .7.

(3) If, in the first test, vision is less than 1.0, and if, in the second test, vision is impaired by the convex lens, then next replace the convex lens by the − .75 D. lens. If vision is now improved so that a smaller line is read, say the second or the first, then the eye is myopic, and may be recorded thus; V. R. E. = .7 + My. = .9, or V. R. E. = .7 + My. = 1.[1]

(4) Place the solid disk before the right eye, and test the left eye similarly. Record the results for each eye separately, *e.g.*, V. R. E. = 1. Em. V. L. E. = .6 + My. = .9.

[1] This test may be nullified in some instances by a ciliary spasm in a hyperopic eye which may simulate myopia of almost any degree.

B. TEST FOR ASTIGMATISM

Apparatus.—Trial frame and lenses as above. Verhoeff's astigmatic chart (Fig. 30).[1]

Preliminaries.—Place the chart on the wall, and seat S as in the previous test. Be sure that S's head is held squarely erect. If S

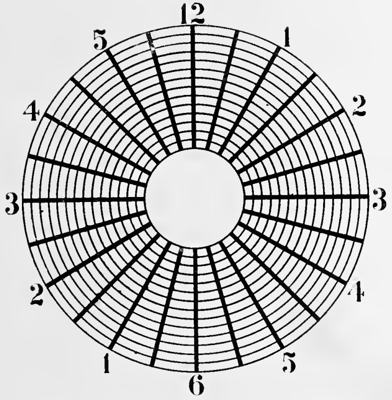


FIG. 30. VERHOEFF'S ASTIGMATIC CHART.
About $\frac{1}{5}$ natural size.

has been found to be myopic or hyperopic, place in the trial frame the lenses which correct, at least partially, this defect.

Method.—Place the solid disk in the frame before the left eye. Ask S whether one or more of the radiating lines seem to him sharper and blacker than those at right angles to them. If he answers in the affirmative, astigmatism is present. This result may be confirmed by causing S to move his head from one shoulder to the other, in which case the location of the sharpest lines should shift in a corresponding manner. The degree of astigmatism may be

[1] Any standard astigmatic cnart may be substituted, but Verhoeff's is, in the author's judgment, best adapted both for making evident the presence of astigmatism and for determining approximately its axis.

roughly judged by the positiveness and readiness of *S*'s answer; its axis may be determined approximately by his designation of the blackest line or lines.

Place the disc before the right, and test similarly the left eye.

Since astigmatism may exist either alone or in combination with some form of ametropia, it should, when found, be recorded with the previous determination, *e.g.*, V. R. E. = .7 + My. = .9 + As.

If vision is less than .7, but no form of ametropia can be demonstrated, the defect is recorded as amblyopia, *e.g.*, V. L. E. = .6 + Am.

To summarize the two tests: emmetropia is indicated (unless strain symptoms point to concealed hyperopia) by the reading of the smallest line and subsequent blurring by the convex lens, hyperopia by improvement or lack of impairment of vision by the plus lens, myopia by vision less than 1., which is improved by the concave lens, astigmatism by unequal clearness of the radiating lines, amblyopia by vision less than .7 without demonstrable refractive error.

Results.[1]—(1) The *frequency* of defective visual acuity is somewhat difficult to state accurately owing to the differences in method and in degree of rigor and precision that have characterized the many investigations upon this point. In especial, a great many investigations in school systems have been made by simple distance tests without the aid of lenses, so that hyperopia, the most frequent defect, has gone practically unmeasured. The general outcome of these simple tests is quite uniform, viz: that one child in three in the public schools suffers from visual defect. Typical figures are those obtained by Welch (20) at Passaic, and Smedley (15) at Chicago; the latter reports that 32 per cent of the 2030 boys and 37 per cent of the 2735 girls examined were defective in vision. While more than half of these defects are of a minor degree, yet, as already indicated, these may be productive of immediate distress and entail serious consequences if neglected.

[1] For a general discussion of the examination of eyesight, with special reference to the eyesight of school children, consult Barry (1) Calhoun (3), Carter (4), Cohn (5), Gould (6,7), Hope and Browne (8), Kotelmann (9), Newsholme (10), Risley (12,13), Schmidt-Rimpler (14), Snell (16), Stilling (17) and Young (21).

On the other hand, examinations that have been conducted by skilled opthalmologists with some refinement of method indicate a much larger percentage of defect.[1] Risley's figures indicate that it would be more correct to state that seven children in eight, than that one in three, are ametropic. As chairman of the Philadelphia committee that examined some 2500 children, he gives the refraction at $8\frac{1}{2}$ years of age as hyperopic in 88.11 per cent of cases, emmetropic in 7.01 per cent, and myopic in 4.27 per cent: at 17.5 years as hyperopic in 66.84 per cent, emmetropic in 12.28 per cent and myopic in 19.33 per cent.

(2) From the above figures we may conclude that the *eye in early childhood* is an incomplete eye, naturally underfocussed and poorly adapted for near work. But, as general bodily maturity approaches, the eye under optimal conditions tends to become emmetropic. Conditions of modern life, however, are not optimal for the eye, but rather encourage overuse and neglect, with the consequence that these, especially when added to astigmatism or other congenital defects, produce eye-strain, myopia, and other disturbances of vision.

(3) Cohn, especially, has demonstrated that *myopia* is essentially a disease of civilization and culture; that it is infrequent in peasants and those who lead an outdoor life, and progressively more prevalent and of higher degree as persistent study and near work continue. Thus, in gymnasia, he found that the percentage of myopia increased during six years of study in the following manner: 12.5, 18.2, 23.7, 31.0, 41.3, 55.8. Similarly, of 138 pupils at the Friedrich gymnasium who were examined twice at an interval of 18 months, he found at the first test 70 normal and 54 myopic, whereas in the second test, 14 of the 70 had become myopic, 28 of the original myopes had developed a higher degree, and in 10 per cent, serious structural changes had taken place in the retina.[2]

[1] Indeed, there are specialists who assert that an absolutely perfect pair of eyes does not exist.

[2] Dürr, in explanation of the alarming prevalence of myopia in Germany as contrasted with other countries, has sought to show its dependence upon the excessive demands of the German school system: he estimates that, during the years 10 to 19, the typical English boy spends in study 16,500 hours, in exercise 4500; the French boy, in study 19,000, in exercise 1300; and the German boy, in study 20,000, and in exercise 650. These statistics showing the relation between myopia and excessive near work could be multiplied almost indefinitely.

(4) Smedley, on the basis of his method of correlation by grades, asserts that "a smaller per cent of the *pupils at and above grade* have defective sight than those below grade."

(5) Smedley further demonstrates that defective vision is extremely common in *backward and troublesome children,* and that this fact may be a partial explanation of the behavior of such children. Thus, according to the tests employed at Chicago, 48 per cent of the boys of the John Worthy School were subnormal in visual acuity, as contrasted with 28 per cent of the boys of the same average age in other schools. Moreover, "many of the John Worthy boys had strabismus, hypermetropia, and astigmatism, conditions which would induce asthenopia when the eyes were used in close and long application to books, and it is easy to believe," adds Smedley, "that the strain thus set up when an attempt was made to study was a factor in producing dislike for school and subsequent truancy."

(6) Van Biervliet (19) has sought to obtain a correlation between the visual acuity test and *intelligence,* not by direct reference to visual acuity itself, but to the mean variation measured by a series of tests of acuity, *i.e.*, to what he terms the capacity of attention. In brief, the method was to compute a fraction, of which the denominator represented the average distance at which the test was visible, and the numerator the mean variation of the several trials. As measured by this arbitrary index, the 10 brightest and the 10 dullest of 300 university students were related, in terms of a common denominator, as 19/1000 and 62.5/1000. Binet (2), however, points out that the dull students had the better eyesight, *i.e.*, the larger denominator, and suggests that the index be taken directly from the mean variation.

(7) *Right-eyedness.* While both eyes are employed for binocular vision, there is some evidence that most persons 'favor' one eye, whenever, for any reason, binocular vision is not in use, *e.g.*, in looking through a microscope or telescope. Van Biervliet (18) has measured the visual acuity of 100 persons, whose optical defects had previously been corrected, with the result that the favored eye very uniformly excels the unfavored one in visual acuity by one-ninth: he further asserts that right-handed people are right-eyed and left-handed people left-eyed, and that the same

sort of sensorial asymmetry can be demonstrated in audition, cutaneous discrimination, and discrimination of lifted weights.

NOTES.—Statistics of visual defect are rendered difficult of comparison, not only by differences in the methods followed in the investigations, but also at times by failure to state whether examinations were made with or without the glasses actually worn by pupils, or, in case such statement is made, to indicate the effect upon the results of including or excluding trials made with these glasses. Cohn distinguishes between visual capacity proper (*Sehleistung*) and visual acuity (*Sehschärfe*), which is the efficiency when proper glasses are used. But since numbers of children are daily wearing improperly fitted glasses, one almost needs another term to indicate the vision that is had with these glasses.

The focal strength of lenses was formerly indicated in the English, or inch system, but it is now more common among opticians to make use of the metric system, in which one dioptric (D) represents a lens whose focal distance is one meter, 2 D. a lens twice as strong, *i.e.*, with a focal distance of 50 cm., etc.

REFERENCES

(1) W. F. Barry, The hygiene of the schoolroom, N. Y., 2d ed., 1904. Pp. 191.

(2) A. Binet, À propos de la mesure de l'intelligence, in A. P., 11: 1904 (1905), 69–82.

(3) A. M. Calhoun, Effects of student life upon eyesight. Bureau Educ., Washington, 1881. Pp. 29.

(4) R. B. Carter, Report on vision of children attending elementary schools in London, London, 1896. Pp. 16.

(5) H. Cohn, (*a*) The hygiene of the eye, Eng. tr., London, 1886. (*b*) Die Sehleistung von 50,000 Breslauer Schulkindern, nebst Anleitung zur ähnlicher Untersuchungen für Aerzte u. Lehrer, Breslau, 1899. Pp. 148. (*c*) Was haben die Augenärzte für die Schulhygiene geleistet und was müssen sie noch leisten?, Berlin, 1904. Pp. 35. (Contains bibliography of 74 titles.)

(6) G. M. Gould, The cause, nature and consequences of eyestrain, in Pop. Sci. Mo., 67: 1905, 738–747.

(7) G. M. Gould, Biographic clinics, 5 vols., Phil., 1903–7.

(8) E. Hope and E. Browne, A manual of school hygiene, Cambridge, Eng., 1904. Pp. 207.

(9) L. Kotelmann, School hygiene, Eng. tr., Syracuse, 1899. Pp. 382.

(10) A. Newsholme, School hygiene, Boston, 1894. Pp. 140.

(11) Rept. com.on statistics of defective sight and hearing of public school children, in U. S., 1902, ii., 2143–2155.

(12) S. D. Risley, Weak eyes in the public schools of Philadelphia: rept. com. on examination of the eyes of the children in the public schools of Phil., in Phil. Med. Times, 11: 1880-1, 673–685.

(13) S. D. Risley, School hygiene, being pp. 353–418 in vol. 2 of W. Norris and C. Oliver, System of diseases of the eye, Phil., 1897.

(14) Schmidt-Rimpler, Die Schulkurzsichtigkeit u. ihre Bekämpfung, Leipzig, 1890.

(15) F. W. Smedley, Rept. dept. child-study and pedag. investigation, in 46th An. Rept. Brd. Educ., Chicago, 1899–1900. Also in U. S., 1902, i., 1095–1115.

(16) S. Snell, Eyesight and school life, Bristol, Eng., 1895. Pp. 70.

(17) Stilling, Die Kurzsichtigkeit, ihre Entstehung und Bedeutung, in Abh. aus d. Gebiete d. pädag. Psych. u. Physiol. 7: No. 3, Berlin, 1903.

(18) J. van Biervliet, L'asymétrie sensorielle, in Bull. Acad. Royale des Sciences, etc., de Belgiques, 34: Série 3: 1897, 326–366.

(19) J. van Biervliet, La mésure de l'intelligence, in Jour. de psych., 1: 1904, 225–235.

(20) G. T. Welch, Report on the examination of the eyes of the public school children of Passaic, N. J., April, 1896. Also summarized in Ref. 11.

(21) A. G. Young, 7th An. Rept. State Brd. Health Maine. Augusta, 1892. Pp. 399. See especially 128–131.

TEST 15

Balance and control of eye-muscles: Heterophoria.—Strictly speaking, the examination of the condition of the eye-muscles is a physiological test, but because this condition affects clearness of vision, it may be included here with other visual tests.

Each eye-ball is supplied with six muscles. By their action in varying combination, the eye is moved freely in its bed, somewhat after the fashion of a ball-and-socket joint. Under normal conditions, the balance and the innervation of these muscles are such that both eyes move in concert, *i.e.*, the eye-movements are automatically coördinated for purposes of single vision and the lines of regard are restricted to movements where a common fixation point is possible. In some individuals, however, there exists more or less 'imbalance', or asymmetry of eye-movement, so that the two eyes fail to 'track,' as it were.

If we consider only the relations of the visual lines to one another and neglect paralytic affections of the muscles, we may distinguish

between latent tendencies toward asymmetry, or *heterophoria*, and actual or manifest asymmetry, *heterotropia* or strabismus. Following the terminology of Stevens (1), we may define the possibilities as follows: *orthophoria* is a tending of the visual lines in parallelism when the determination is made for a point not less than 6 m. distant; heterophoria is a tending of these lines in some other way under the same conditions. Heterophoria may appear (*a*) as *esophoria*, a tending of the lines inward or toward one another, (*b*) as *exophoria*, a tending of the lines outward or away from one another, (*c*) as *hyperphoria*, a tending of the right or of the left visual line in a direction above its fellow,[1] or (*d*) as tendencies in oblique directions, viz: hyperesophoria and hyperexophoria.

The tendencies just described are tendencies only, and are latent or concealed in the ordinary use of the eyes on account of the strong 'desire' for binocular vision. For their discovery, accordingly, it is necessary to resort to means for eliminating, so far as possible, this reflex or automatic correction of the latent tendency. The means most commonly employed, as illustrated in the tests that follow, is the establishment of disparate images on the two retinas.[2]

When binocular vision is not habitually attained, the tendencies above described are no longer latent, but manifest, and heterotropia (strabismus or squint) is the result. Heterotropia may appear as esotropia, converging strabismus, or deviation of the visual lines inward; as exotropia, diverging strabismus, or deviation outward; as hypertropia, strabismus sursumvergens or deorsumvergens; or as compound deviations, termed by Stevens hyperesotropia and hyperexotropia.

The most obvious immediate result of heterotropia is diplopia or double vision, a very annoying, but not usually a permanent symptom, because the person thus affected soon comes to neglect the bothersome image from the 'squinting' eye, and to take account only of that from the 'fixing' eye. In time, there results, usually, a limitation of the movements and of the retinal

---

[1] The term does not imply that the line which is too high is at fault, but merely that it is higher. Hence, of course, the lack of necessity for any term to indicate that one line is lower than the other.

[2] This assumption that voluntary attempts at fusion will be renounced if the two images are sufficiently disparate, is not entirely correct, and in so far, it is not always possible to make an accurate determination of heterophoria, particularly when slight, by means of the principle of diplopia. Slight heterophoria, moreover, is not to be regarded as abnormal.

sensitivity of the squinting eye (exanopsic amblyopia), which is one of the most interesting instances of the loss of function through disuse.[1]

Strabismus or heterophoria is functionally associated with ametropia; in particular, divergent displacement is more apt to be associated with myopia, and convergent displacement with hyperopia, probably as a consequence of the straining after clear vision under the hyperopic handicap.

The chief instruments for the detection of muscular asymmetries are prisms of varying construction, the Maddox rod, and the stenopaic lens. Stevens' phorometer is a device for holding and rotating prisms with accuracy and under optimal conditions. The phoro-optometer is a combination of the phorometer with other instruments, such as the Maddox rod, Risley's prism, etc.

Two tests are here detailed, the Maddox rod and the prism test. Both are convenient, portable, and inexpensive, but possess the disadvantage common to all tests for heterophoria held close to the eye, viz: that *S* does not always completely renounce the fusion-impulse.

In the Maddox test, the so-called 'rod' transforms for one eye the flame of a candle into a long narrow streak of red light, while the other eye sees the candle flame naturally. Heterophoria is indicated by the lack of coincidence in these two images.

The prism test, which is essentially an auxiliary test, consists in producing artificial displacement of images by means of the prisms, and measuring *S*'s ability to produce voluntary fusion of these displaced images.

A. THE MADDOX ROD TEST

Apparatus.—Maddox multiple red rod (Fig. 31). Trial frame (Fig. 29), Candle. [A set of trial prisms may be added.]

Preliminaries.—Place the lighted candle on a level with *S*'s eyes and 6 m. distance. Adjust the trial frame.

Method.—(1) Let *S* close his left eye: place the Maddox rod in the frame before the right eye with the bars set horizontally. *S* should then perceive a long, narrow, vertical streak of red light. Then let *S* open his left eye and at once state whether the red streak passes exactly through the candle flame.

[1] As this is particularly to be feared in tne case of children, whose eyes have not reacned functional maturity, prompt medical attention to strabismus is highly imperative.

(2) Turn the rod until the bars run vertically. *S* will see a horizontal red streak. Let him open his left eye and at once state whether the streak passes exactly through the candle flame.

RESULTS.—In the first test, the possible results are: (*a*) the line passes through the flame, orthophoria (Fig. 32), (*b*) the line passes to the right of the flame, esophoria or homonymous displacement (Fig. 33), (*c*) the line passes to the left of the flame, exophoria or crossed displacement (Fig. 34).

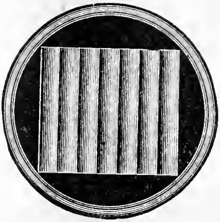


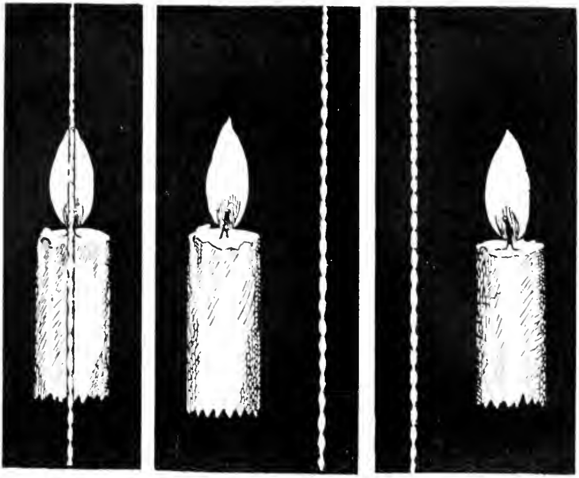
FIG. 31. MADDOX MULTIPLE ROD.

In the second test, the possible results are: (*a*) the line passes through the flame, orthophoria (Fig. 35), (*b*) the line passes below the flame, right hyperphoria (Fig. 36), (*c*) the line passes above the flame, left hyperphoria (Fig. 37).

NOTES.—Next to orthophoria, esophoria is the most common condition. Unequal vertical adjustment, hyperphoria, is not common, save that an upward deviation of the squinting eye is almost always associated with high degrees of convergent strabismus.

If the latent asymmetry is but slight, there may appear a more or less rapid corrective movement: *S* will then notice lack of coincidence of the line and the flame when the left eye is opened, but the two images soon fuse together. On the other hand, if the asymmetry is larger, *E* may determine its degree by placing prisms before the left eye and ascertaining by trial how strong a prism is needed to enable fusion to occur.

If both horizontal and vertical imbalance is observed, the defect is hyperesophoria or hyperexophoria. This may be demonstrated, if desired, by placing the Maddox rod in an oblique position.



FIGS. 32–34. ILLUSTRATING ORTHOPHORIA, ESOPHORIA, AND EXOPHORIA, RESPECTIVELY.

As revealed by the Maddox rod when used before the right eye for horizontal deviation. (De Schweinitz and Randall).

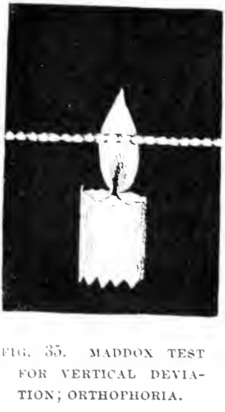


FIG. 35. MADDOX TEST FOR VERTICAL DEVIATION; ORTHOPHORIA.

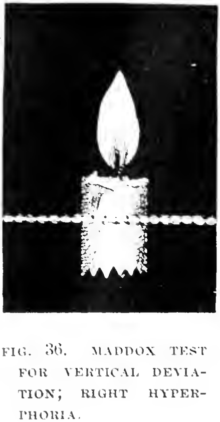


FIG. 36. MADDOX TEST FOR VERTICAL DEVIATION; RIGHT HYPERPHORIA.

FIG. 37. MADDOX TEST FOR VERTICAL DEVIATION; LEFT HYPERPHORIA.

Stevens's stenopaic lens (Fig. 38) may be substituted for the Maddox rod. A single determination then suffices for both horizontal and vertical displacement. In orthophoria, the candle flame appears in the center of a diffused disc of light; in hetero-

FIG. 38. STEVENS' STENOPAIC LENS.

phoria, it is displaced to the right or left, above or below, or obliquely, in a manner corresponding to that of the Maddox line-and-flame test (Fig. 39). The stenopaic lens consists of a convex lens

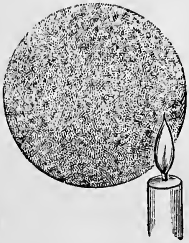


FIG. 39. HETEROPHORIA, AS REVEALED BY THE STEVENS LENS.

of 13 D., covered, save for a very small opening in the center. The principle is again that of disparate images.

B. TEST WITH PRISMS

APPARATUS.—Trial frame. One 2-degree, one 8-degree, and two 20-degree prisms, of the circular pattern for the trial frame.[1] Candle.

[1] These prisms permit *E* to test *S*'s ability to overcome the degrees of displacement that are considered standard for the three positions: their cost is about $4. For $10, however, a fairly complete set of prisms may be purchased, which will permit a more flexible test.

PRELIMINARIES.—Place the lighted candle on a level with *S*'s eyes and 6 m. distant. Adjust the trial frame.

METHOD.—(1) To test abduction, or *S*'s ability to overcome a standard amount of displacement by rotating the eyes outward, place the 8-degree prism before one eye with the base in, *i.e.*, toward the nose.

(2) To test *S*'s ability in adduction, or forcible convergence, place a 20-degree prism, with the base out, *before each eye.*[1]

(3) To test *S*'s ability in sursumduction (compensation for vertical displacement), place the 2-degree prism, with the base either up or down, before one eye.

RESULTS.—With orthophoria, *S* should secure fusion under the conditions imposed, if not at the first trial, at least after a few trials on different days. Failure to accomplish this, or ability to overcome larger angular displacements than those cited, is indicative of heterophoria, or of other inequalities in the set of the eye-balls, *e.g.*, declination.[2]

NOTES.—This test may, of course, be applied to cases in which either orthophoria or heterophoria is present. It may be of value in measuring *S*'s control of his eye-muscles, not only as a matter of optical hygiene, but also in conjunction with tests and experiments of a psychological nature, *e.g.*, stereoscopy, binocular fusion, and visual space-perception in general.

REFERENCES

(1) G. T. Stevens, A treatise on the motor apparatus of the eyes, Phil., 1906. Pp. 496.

(2) W. N. Suter, The refraction and motility of the eye, Phil., 1903. Pp. 390.

TEST 16

Color-blindness.—This test continues the examination of the functional efficiency of the eye as a sense-organ. It has obvious practical import, as well as high theoretical significance in connection with the theory of vision.

[1] The ability to overcome prisms by convergence is about 50 degrees, according to Stevens, but an exact standard cannot be stated.

[2] For further details of this, and otner forms of prism test, consult Suter and Stevens.

The retina of the normal eye is not equally sensitive to color stimuli in all portions. On the contrary, simple experiment[1] will demonstrate the existence of three distinct zones: an inner, efficient zone, over which we see all colors; a middle zone, over which we see only blues, yellows, blacks, whites, and grays, and an outer, totally color-blind zone, over which we see nothing but blacks, whites, and grays.

Color-blindness may be regarded as an arrest of development, or in some cases as a pathological modification, of these normal zones.

The description and classification of forms of color-blindness has been much obscured, if not even actually retarded, by confusion in nomenclature, especially seen in the adoption of terms based on unwarranted, preconceived theoretical notions as to the nature of color vision.

If we follow the clue afforded by the distribution of color sensitivity on the normal retina, we should expect to encounter total color-blindness, or partial color-blindness of a red-green type. Blue-yellow blindness would seem theoretically unjustifiable, at least as an arrest of development. And in fact we do find that all recorded cases of so-called blue-yellow (or violet) blindness are of doubtful character, and at least pathological in nature.

Total color-blindness is well authenticated, but rare. Here, too, the defect is pathological, and is accompanied by a reduction in visual acuity, by nystagmus, and other serious disorders of the visual organ.

We are left, therefore, with red-green blindness as the typical and characteristic form of color deficiency. As revealed in tests, this defect consists in the inability correctly to distinguish certain tones, particularly unsaturated tones, of red and green. The colors thus affected are invariably in pairs, *i.e.*, the individual who fails to perceive correctly a given green will also fail to perceive correctly its complementary red, and conversely.

In actual vision, certain reds and certain greens appear neutral or gray,[2] while tones in which red and green are conjoined with

[1] For details, consult Titchener (12, Part I., Section 9).

[2] The green which appears as gray is a somewhat bluish green, lying between the *b*-line and the *F*-line of the spectrum, and having a wave length between 500 $\mu\mu$ and 490 $\mu\mu$: the complementary red is a non-spectral purplish red.

blue or yellow are seen as bluish or yellowish. The spectrum is thus divided into a long-waved yellow and a short-waved blue section. By dint of daily experience, however, the color-blind individual develops a capacity to recognize some reds and greens by means of secondary criteria, such as brightness (tint) and saturation (chroma), and familiarity with the application of color nomenclature (grass is green, cranberries red, etc.), so that the defect may exist unrecognized, either by himself or by his acquaintances, until chance compels the recognition or discrimination of tones to which these criteria can not be applied. Hence arises the necessity, in the administration of tests, of displaying a large number of colors of varied saturation and brightness, in order that, for any individual, some combination or series of combinations of colors may be found, in the recognition of which these criteria can not be used. Here, too, appears in part the explanation of the seeming individuality of the defect.

Despite, and quite apart from, this variability, however, two very definite sub-types of red-green blinds may be distinguished, though the distinction is of theoretical rather than of practical significance: (*a*) those who locate the brightest part of the spectrum, as do normal persons, in the yellow, and (*b*) those who locate the brightest part of the spectrum in the yellow-green region of the spectrum. For the latter, the entire blue end of the spectrum appears relatively brighter than to the normal eye. This type is relatively infrequent. Individuals who belong to the former type are called by von Kries (13) *deuteranopes*, and are erroneously called by many writers green-blinds: those who belong to the latter type are called by von Kries *protanopes*, and erroneously red-blinds.

This confusing, if not erroneous, terminology is to be referred to the Helmholtz theory of color vision, wherein the three primary visual and retinal elements are assumed to be red, green, and violet. In theory, on this basis, it is evident that an eye might possess all three, or but two, or but one, of these visual elements; that, in other words, an eye might possess trichromate, dichromate or monochromate vision. Ordinary red-green blindness would, therefore, on this theory, be a form of dichromasy with the deficiency actually due to loss either of the red or of the green element.

This terminology ignores the fact that color-blindness invariably goes in pairs,[1] but the terminology persists and is current in nearly all popular discussions of the topic.

This forms one of the most conspicuous instances of the imposition of theoretical convictions upon the interpretation, and even upon the observation of facts. Not a few discussions of color-blindness that make pretence to scientific accuracy exhibit this error, and, one may add, other errors of a more inexcusable sort. The reader may consult, for instances, a book by Abney (1), which embodies his Tyndall lectures of 1894, and a magazine article by Ayers (2). In Abney there will be found a colored frontispiece, taken from the Report of the British Association Committee on Color Vision in 1892, which purports to show the spectrum as seen by the color-blind. The spectrum is shown in green and blue: what becomes of yellow is not explained. In Ayers's article there will be found some very pretty colored pictures of roses and Venetian scenes as observed by the color-blind,—pictures that are good examples of the illustrator's art, but absolutely false examples of color vision. Mrs. C. L. Franklin (6) has charitably applied the term "pseudo-scientific" to such writing. A more nearly correct representation of the spectrum seen by the color-blind is given by Thomson (11).

Holmgren contrasted total color-blindness with partial color-blindness, and divided the latter into complete partial color-blindness and incomplete partial color-blindness (confusions with the green, but not with the red, test skein). This division has not been often used, but the term 'color-weakness' has been extensively employed in place of Holmgren's incomplete partial color-blindness, though not quite correctly, because this group, as determined by the Holmgren test, may embrace both von Kries's deuteranopes and the so-called color-weak.

The most recent and authoritative study of *color weakness* has been conducted by Nagel (9). According to his investigations, the color-weak, with the rarest exceptions, are (in Helmholtzian terminology, after König) *anomalous trichromates*, *i.e.*, they possess all the elements of color vision, but exhibit certain anomalies, of which the following are most prominent: (*a*) a considerable reduction of sensitivity and discriminative capacity in the region of yellow-green and green, (*b*) inability to recognize colors, particularly red and green, when of reduced intensity, small area, or brief

[1] There have been reported a few exceptions to this rule, which are difficult of explanation by any theory of color-vision.

exposure, (*c*) rapid retinal fatigue to colored stimuli, (*d*) slow recognition of color tones, amounting, for reds and greens, to 20–50 fold the normal time, (*e*) increased dependence upon brightness differences, (*f*) very marked augmentation of simultaneous and successive contrast.[1]

Nagel suggests a further subdivision of anomalous trichromasy into protanomalous trichrómasy (a lessened excitability to red corresponding to protanopia) and deuteranomalous trichromasy (corresponding to deuteranopia). But, as he points out, this distinction may be somewhat premature, since the facts are not yet clearly established, and since a case has recently been discovered in which anomalous trichromasy and dichromasy appear to coexist in the same eye.

In theory, the color-weak are not to be identified with the color-blind. Their defect ranges all the way from forms which are to be distinguished from normal vision only by careful tests to forms which closely approximate true dichromasy or color-blindness. For practical purposes, however, they must be identified with the color-blind, because they are incapable of making those color discriminations that the conditions of railway and marine service demand.

It is, perhaps, true that some disasters may be traceable to this defect in color vision, which has escaped the detection of medical examiners who have used only the standard wool tests. Thus, in Germany, among 1778 members of railway regiments, all of whom had passed the wool test and many of whom had also passed Stilling's test, 13 dichromates and 31 anomalous trichromates of various types were discovered by the use of Nagel's test in the hands of military physicians. Baird, however, contends that statistics of railway accidents show no trace of this factor.

Color-blindness may be binocular or monocular. The latter is rare, but naturally of great theoretical importance in determining the nature of color-blindness.

Color-blindness is usually congenital, and then incurable. The common form, red-green blindness, is to be regarded as an arrest of development, a reduction from normal trichromasy, or reversion to a more primitive form of retina. All acquired cases, variously attributed to traumatism, neuritis, atrophy of the optic nerve, hysteria, excessive fatigue, over-indulgence in tobacco, are accompanied by lessened visual acuity, are pathological, and of relatively small concern to the theory of color-vision.[2]

---

[1] For an excellent discussion of color weakness and the use of tests, the reader is advised to consult J. Rosmanit (10).

[2] For illustrative cases, consult Collin and Nagel (4).

Color-blindness seems to have been first noted in literature in 1684, but first described accurately by Dalton, the celebrated English chemist, in 1794. The first attempt at a systematic examination of a large number of cases was made by Seebeck in Berlin in 1837 by the aid of colored papers. The first systematic examination of railway employees dates from 1875, when a serious accident in Sweden led Holmgren of the University of Upsala to devise his well-known wool test and to induce officials to adopt it.[1]

The chief devices and methods for testing color-blindness are Holmgren's, Galton's, Thomson's, Oliver's, and other assortments of colored worsteds, Stilling's pseudo-isochromatic charts, Nagel's card test, spectroscopic examination, various contrast tests, and the use of equations of mixed colors, particularly Nagel's equation-apparatus, and Hering's apparatus,[2] which enables the examiner to adjust a color equation of transmitted light that shall appear to the color-blind as uniform gray. Nagel's or Hering's apparatus is to be recommended for careful psychological tests. In addition, numerous forms of color-blindness lantern (Williams', Friedenberg's, Oliver's, etc.) have been devised for testing railroad and marine employees by simulating the conditions of night-signalling, and soiled signal-flags have been used for similar purposes, while Henmon has proposed a discrimination-time test.

Two forms of test are here described: the familiar and widely used Holmgren wool test, adopted by the American Ophthalmological Society, and Nagel's new card test, which has now been specially revised for the diagnosis of color-weakness and of other variant types of defect. Both of these tests are inexpensive, compact, and portable. They may be employed in conjunction with one another.

A. THE HOLMGREN WOOL TEST

MATERIAL.—Holmgren's worsteds. Sheet of light gray or white cardboard or a similarly colored cloth.

METHOD.—(*a*) *Full procedure.* (1) Remove the three large test skeins, pale green, red and rose, Nos. 101, 102, 103. Scatter the remaining skeins over the cloth or paper in *diffuse daylight*

[1] For other details of the history of color-blindness as well as a discussion of methods, though not brought down to date, consult Jennings (8) and Thomson (11).

[2] See Titchener (12, Pt. II, p. 7).

only.[1] Hand to S the green test skein, No. 101, and direct him to pick from the table all those skeins that resemble the test skein *i.e.*, all the tints and shades of that color. Explain that there are no two specimens alike, and that an exact match is not required. It will do no harm to illustrate the process by selecting two or three skeins for him, provided these are afterward mixed with the pile. To save time in explanation, other S's may be allowed to watch this demonstration.

(2) If hesitation appears, or if grays, browns or reds as well as greens are selected, continue the test by use of the rose skein, No. 102. The typical color-blind will then select some blues or purples, or, less often, grays or greens.

(3) Finally, the red test skein, No. 103, may be used, though many red-green blinds have little difficulty with this test on account of the strong saturation of the test skein.

In all three tests, preserve a careful record of the skeins selected by S's who deviate in any particular from the normal.

(*b*) *Abbreviated procedure.* This test may be used for quick preliminary examination. Place irregularly on the cloth four green skeins (*e.g.*, 5, 17, 19, 49) and eight 'confusion' skeins of gray, brown, and pink (*e.g.* 8, 12, 27, 40, 44, 46, 54, 60). Hand to S the pale green standard, No. 101, and require him to pick from the cloth as rapidly as possible four skeins that match the test skein (in the sense previously described). Allow him approximately 4 sec. to make this selection. If this test can not be promptly and accurately executed, examine S further by the full procedure.

TYPICAL RESULTS.—(1) About 4 per cent, of men and less than 0.5 per cent of women[2] are color-blind: the most common defect is red-green blindness of the form known as deuteranopia.

(2) The following are actual selections of a typical red-green blind. By assembling these skeins E can gain the best idea of the nature of the confusions likely to be discovered.

[1] When not in use the skeins must be kept carefully enclosed in their box, as they will fade or change color if continuously exposed to light.

[2] Recent tests by Dr. S. P. Hayes seem to indicate that the frequency of color defect in women is much greater than this.

Green standard: 5, 8, 10, 12, 13, 19, 32, 38, 45, 47, 53, 60, A, C, D, K, N. (Occasionally some pink, like 27, is also selected.)

Rose standard: 22 to 28, 30, 71, 72, 74, 76, 78, 79, 80, 82, 84, 86, 88, 90, 91, 92, 93.

Red standard: 21, 29, 31, 33, 35, 37, 39, 73, 75, 77, 85, 87, 89.

NOTES.—Inability to name colors rightly has sometimes been erroneously mistaken for color-blindness, but the term must be applied only to instances of actual inability to *see* colors rightly. Consequently, no color-blindness test should *hinge* upon the ability to name correctly the various colors presented, and, in the conduct of the Holmgren test at least, reference to color names should be avoided if possible.

If *S* works very slowly and hesitatingly, but finally makes a correct selection, this may indicate several possibilities, which should be tested by further study of the case. (*a*) The slowness may be due merely to extreme cautiousness on his part, coupled with some anxiety or uneasiness about the test, or with failure to understand clearly just what is wanted. (*b*) The slowness may be due, in the case of very young children or untutored adults, to gross ignorance of, and unfamiliarity with colors. (*c*) The slowness may be indicative of color-weakness, in which case Nagel's test should be applied for further diagnosis of the defect.

All instances in which specific color differences are at first recognized with difficulty or not at all, but in which, after coaching or instruction, an efficiency is developed adequate for passing the test in use, must be looked upon with suspicion, and it must not be assumed forthwith that color-blindness has been cured by training, for either the cautiousness or ignorance just mentioned were present at first and removed by the training, or the conditions of the test were too simple, and secondary criteria were developed by *S*.

If apparent cases of blue-yellow, or of total color-blindness are discovered, these should, if possible, be given most careful examination by an expert psychologist.

It is obvious that many callings are, or should be, closed to the color-blind, *e.g.*, railroading, marine and naval service, medicine, chemical analysis, painting and decorating, certain branches of botany, microscopy, mineralogy, the handling of dry goods, mil-

linery, etc. In some phases of school work, the color-blind pupil is likewise at an evident disadvantage. The test should, accordingly, be regularly instituted in the early years of school life, in order that the existence of the defect may be made known to the child as soon as possible.

B. THE NAGEL CARD TEST

MATERIAL.—Nagel's color-blindness cards, 4th edition.[1]

Main Test, Part I.

METHOD.—Spread out the cards of Section A upon a table, in chance arrangement and in good daylight illumination. These must be observed by *S* from a distance of *75 cm.*, never less.

(1) Ask *S* to point out a card upon which *red or reddish spots* are seen (do not suggest bluish red or rose). Ask for a second and a third such card.

Normal and color-weak *S*'s will answer correctly; color-blind *S*'s will usually select a card that contains yellow-green and no red, *e.g.*, No. 6 or 11.

(2) Ask *S* to point out a card that contains *red spots only.*

Normal *S*'s answer correctly; color-weak *S*'s usually answer correctly, but may select a red and gray card; color-blind *S*'s frequently select No. 12, *i.e.*, yellow-green and yellow-brown.

(3) Ask *S* to point out cards that contain (*a*) *green spots only,* and (*b*) *gray spots only.*

Normal *S*'s will select No. 5 (for green) and No. 9 (for gray), but No. 16 may be allowed. Normal *S*'s may at first also select cards which contain combinations of gray and green, such as Nos. 4, 13, 14, but will reject these after more careful examination. Color-weak *S*'s will more often select these last mentioned cards and not correct the mistake. The color-blind are helpless, and can select the correct cards only by mere chance.

[1] Nagel's test can not be purchased by the general public in the open market, but may be imported for use by physicians, psychologists, and biologists, or for institutions represented by such specialists. Importation can be arranged through C. H. Stoelting Co., N. Green St., Chicago.

Main Test, Part II.

The object is to extend the previous test and to give more careful attention to possible cases of anomalous trichromasy.

METHOD.—Display the first three cards of Section B (B1, B2, B3) at a distance of *30 cm.*

(4) Point out Card B1 (green and brown) and ask *S*: "What colors do you see here?"

Normal *S*'s will at once note the presence of two colors, though they may not name them correctly as green and brown, but may use such terms as dark yellow, gray, yellowish. Color-weak *S*'s may answer similarly, but with less certainty, and frequently call all the spots green or all brown, or use the terms green and red. Color-blind *S*'s call all the spots red, yellow, brown, or green, and frequently call special attention to differences in brightness of the spots.

(5) Point out Card B2 or B3 (red and brown) and again ask *S*. "What colors do you see here?"

Normal *S*'s answer correctly, though they may call the brown, dark-yellow. (If hesitation appears over this brown, *E* may inquire whether the color might be yellow, or gray, or green, or brown.) Color-weak *S*'s call this card red and green (or greenish or olive). Color-blind *S*'s answer as in (4).

Supplementary Test A

This is designed for testing all difficult or obscure cases (unintelligent or dull *S*'s, color-weaks, etc.) that have not given clear results by the preceding tests. It should be administered to all *S*'s that have shown hesitation in the preceding trials, particularly when green was under test.

METHOD.—(6) Display the cards of Section A. Ask *S* to point out *all* the cards upon which red or reddish spots are seen.

(7) Ask similarly for all cards on which green spots are seen.

(8) Ask what cards contain *no* green spots.

(9) Display Cards B1 and B2; ask *S* to point out the red and green spots with a pencil.

If Tests 6–9 are answered correctly, *S* is certainly not color-blind.

(10) If doubt remains as to whether *S* is color-weak, display in rapid succession several cards from Section A, and insert in the series the four cards of Section B. Ask *S* to state upon which cards green spots are found. If *S* mentions B2 or B3, he is color-weak, *i.e.*, an anomalous trichromate.

Supplementary Test B

This is designed for the special diagnosis of protanopia and deuteranopia for scientific and statistical purposes.

Method.—(11) Display Card B4, and direct *S*'s attention to the pairs of spots indicated by asterisks. Ask *S* which spot in each pair is darker. If the green is selected, *S* is 'green-blind' or deuteranopic, *i.e.*, the relative brightness is approximately that of normal vision: if red is selected, *S* is 'red-blind,' or protanopic. Red-anomalous and green-anomalous types may be similarly differentiated. Red-blinds will also find the red spots decidedly darker than the brown on Card B3.

(12) Very rare cases of blue-yellow blindness may be recognized by inability to distinguish the blue-green from the yellow-green spots on Card A6 and by the designation of all the spots on B2 as red.

Notes.—To use Nagel's test successfully, it is imperative that the following cautions be observed. (1) The directions for conducting the main tests must be strictly followed before attempting any supplementary or variant tests. (2) *E* must adopt a quiet, sympathetic manner, free from any sign of irritation or impatience. especially when dealing with slow or stupid *S*'s, or even with those who are plainly attempting deceit. (3) During the test, *E* must carefully avoid informing *S*, whether directly or by suggestion, of any mistakes he may make. Discussion or criticism of *S*'s selections is out of place. For the sake of future tests, it would be desirable not to explain *S*'s errors to him even after the test.

To avoid the effect of possible collusions between *S* and previous *S*'s, Tests 4 and 5 may be preceded by similar questions applied to several cards taken from Section A.

If *S* has decidedly low visual acuity, this must be corrected, at least approximately, by appropriate lenses, before the color-blindness test is begun.

REFERENCES

(1) W. Abney, Color vision, London, 1895.

(2) E. Ayers, Color-blindness, with special reference to art and artists, in the Century Mag., 73: 1907, 876–889.

(3) J. W. Baird, The problems of color-blindness, in P. B., 5: 1908, 294–300.

(4) Collin and W. Nagel, Erworbene Tritanopie, in Zeits. f. Sinnesphysiol. 41: 1906, 74–88.

(5) Dalton, Extraordinary facts relating to the vision of colors, in Trans. of the Lit. and Philos. Soc. of Manchester, 1794.

(6) Mrs. C. L. Franklin, Magazine Science, in Science, n.s. 25: 1907, 746.

(7) V. Henmon, The detection of color-blindness, in Jour. Phil. Psy. and Sci. Method, 3: 1906, 341–4.

(8) J. E. Jennings, Color-vision and color-blindness, Phil., 1896. Pp. 111.

(9) W. Nagel, (*a*) Fortgesetzte Untersuchungen zur Symptomatologie u. Diagnostik der angeborenon Störungen des Farbensinns, in Zeits. f. Sinnesphysiol., 41: 1906, 239–282, 319–337. (*b*) Zur Nomenclatur der Farbensinnstörungen, *ibid.*, 42: 1907, 65. (Consult also other articles by the same writer in this periodical.)

(10) J. Rosmanit, Zur Farbensinnprüfung im Eisenbahn- und Marinedienste, Vienna and Leipzig, 1907. Pp. 59.

(11) W. Thomson, Detection of color-blindness, being pp. 315–352 in Norris and Oliver, System of diseases of the eye, vol. ii., Phil., 1897.

(12) E. B. Titchener, Experimental psychology, vol. i., N. Y., 1901.

(13) J. von Kries, Ueber Farbensysteme, in Z. P., 13: 1897, 241–324. (Corrective note, p. 473.)

TEST 17

Discrimination of brightness.—The object of this test is to obtain an index, for comparative purposes, of *S*'s ability to distinguish very small differences in brightness, or more exactly, to determine the smallest difference in brightness that *S* can distinguish under simple experimental conditions. The present test omits consideration of chromatic stimuli, and is confined to the discrimination of brightness, first by the use of reflected, secondly by the use of transmitted light.

Visual discrimination has been studied in the laboratory by many competent investigators, *e.g.*, Ament, Aubert, Bouguer, Helmholtz, Fröbes, Kraepelin, Masson, Merkel, Schirmer, Volkmann, and others. Tests of school children by Gilbert (3) and Spearman (6) have followed simpler methods.

In the laboratory, use has been made of Masson's disk, both by daylight and artificial illumination, of the episkotister, or gray

papers and of shadows. Toulouse (9) proposes solutions of aniline colors in glass receptacles. Gilbert used a series of ten pieces of cloth soaked in a red dye of graded intensity. Investigations that are most comparable with the method here proposed are those of Ament (1), Fröbes (2), and Spearman, all of whom made use of gray papers, and of Gilbert, who examined school children, though with chromatic stimuli.

A. DISCRIMINATION OF GRAYS—REFLECTED LIGHT

Apparatus.—Set of 10 test-cards, each composed of two gray strips, 13 x 40 mm., on a white background, 10 x 10 cm. Exposure frame, fitted with a card-holder which may be rotated through

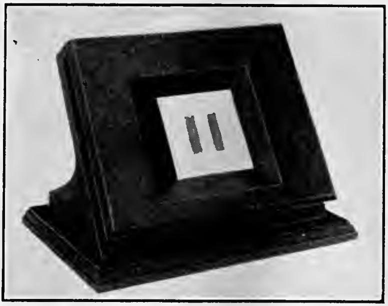


FIG. 40. APPARATUS FOR THE DISCRIMINATION OF GRAYS.

180°, and with a black screen, through an opening (8 x 8 cm.) in which the test-cards may be viewed (Fig. 40). Light gray cloth, about 70 x 160 cm., for a background. Two supports, with angle-pieces, and a horizontal rod 70 cm. long. Headrest.

The cards are numbered from 0 to 9, corresponding to 10 different pairs of stimuli. Each card contains one strip of the lightest or standard gray, and one strip of comparison gray. Card No. 0 represents no difference, or objective equality; Card No. 1 represents the minimal objective difference; Card No. 9 the maximal objective difference and is easily supraliminal for the normal eye. Each card is numbered on the back in such a way that, when looking at the face of the card with the number up, the right strip is the darker; there is also a small black mark on the extreme edge of the card on the side of the darker strip.

The grays used on these cards have been specially prepared, under the author's direction, by S. L. Sheldon, photographer, of Ithaca, N. Y., and have been carefully standardized. Each set of grays is printed from the same negative, on which the original series was formed by graded serial exposures before a sheet of milk glass set in a north window. They will not fade or change their tone, unless brought into contact with chemical fumes or solutions; but, for additional protection, they should be kept under cover when not in use, and never be handled in bright sunlight.

The tones, sizes, and spatial relations of the strips, cards, and background have been selected to eliminate errors that might arise from adaptation and contrast. The size of the strips is slightly smaller than that used by Ament (18 x 45 mm.) and slightly larger than that prescribed by Titchener (10 x 40 mm.) for the demonstration of Weber's law in brightnesses.

PRELIMINARIES.—Place a small table, say 65 x 90 cm., squarely before a window where good diffuse daylight may be secured (preferably a north window with full clear exposure to the sky); leave just enough space between the front of the table and the window for two chairs for *S* and *E*. Spread the gray cloth over the top of the table, and stretch it up vertically at the back edge by means of the supports, so as to form a continuous background of gray, with the vertical back at least 65 cm. high and about 65 cm. distant from *S*'s eyes.

Place the exposure frame in the center of the table at the optimal reading distance (about 35 cm., unless *S* has uncorrected myopia or hyperopia), and adjust its height so that the *top* of the frame is on a level with *S*'s eyes. Adjust the headrest so that *S* may sit erect, squarely before the exposure frame and close to the table-edge, with his back, of course, to the window.

Keep the test-cards conveniently near, but out of *S*'s sight. *E* will find it most convenient to sit at *S*'s right.

METHOD.—(1) Spend 5 min. in giving *S* practise and familiarity with the test. For this purpose, begin with the large-numbered cards, and pass in general toward the smaller numbers, but without following any rigorous order. With each card, rotate the turntable, so that the right strip is now the darker, now the lighter: follow an irregular order, and keep *S* always in ignorance of the actual location of the darker strip, and of the correctness of his judgments. In each trial, *S* must report his judgment *in terms of the right-hand strip,* saying either "darker," "lighter," or "equal.' (Any doubtful cases may be classed as equal.)

When not observing a test-card, S should rest his eyes by directing them toward the gray background. He turns his eyes to the test-card at E's "now," and should be asked to pass a judgment *within 5 sec.* It is not necessary to record results at this point, but from this practise work, S will attain a general familiarity with the test, and E will form a fair idea of S's 'critical' region.

(2) Proceed now, more formally and exactly, to determine S's difference limen by selecting a stimulus difference which has appeared in the preliminary series to be just noticeable for him. Give this stimulus-card 10 times, 5 times with the right strip darker, 5 times with the right strip lighter, but in chance order.[1] Inform S that he will be shown the same card 10 times, but in different positions, of which he is to be ignorant. He may answer "lighter," "darker," or "equal." (Equal judgments may be classed as wrong.) S must not be informed during the series whether his judgments are right or wrong. If S gives 8 right answers in 10, the magnitude of the brightness difference then in use affords the desired index.

(3) Confirm the result by testing S 10 times with a slightly larger difference, and 10 times with a slightly smaller difference. Unless the tests are disturbed by the operation of such factors as fatigue, loss of interest, practise, fluctuations of attention, etc., S will give 9 or 10 correct judgments in the former, and fewer than 8 in the latter test.

Variations of Method.—Test the discriminative capacity of each eye separately, as well as in conjunction. Employ the trial frame of Test 14, placing the solid disk before the untested eye. Care must be taken to avoid visual fatigue under these conditions. This variation of method is of interest in connection with recent work on psychophysical asymmetry and the relations between right-handedness, right-eyedness, right-earedness, etc. (See, for example, Van Biervliet).

If means are at hand to secure effective constant illumination by artificial light, this may be tried for comparison with daylight illumination.

[1] It is convenient to prepare on small slips, beforehand, a number of chance orders, and to follow one of these with each set of 10 trials.

Treatment of Results.—For comparative purposes, S may be ranked in terms of the arbitrary units afforded by the card-numbers. For more exact quantitative expression, however, the results should be expressed in terms of the brightness-differences which correspond to the card-numbers. This correspondence must be worked out by E for the papers employed. Full directions for a simple and sufficiently accurate photometric determination of brightness values of gray papers will be found in Titchener (Pt. I., 35 ff.).

B. DISCRIMINATION OF BRIGHTNESSES—TRANSMITTED LIGHT

Apparatus.—Headrest. Brightness discrimination test (Fig. 41). [This is a box fitted with a high power frosted tungsten lamp, the light of which is reflected from two independently adjustable

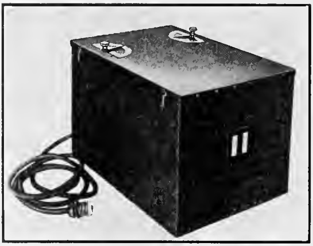


FIG. 41. APPARATUS FOR BRIGHTNESS DISCRIMINATIONS.

white screens upon two oblong, translucent windows, so placed in the face of the box as to give the same dimensions and spatial relations as obtained in the case of the gray strips.]

Preliminaries.—The lamp cord is to be attached to a suitable current (106–110 volts, unless special lamps are ordered). E should endeavor to conduct the test in a dark or darkened room. If a brightly lighted room must be used, the effectiveness of the illumination of the 'windows' may be increased by erecting a protecting screen of cardboard or cloth around them.

Method.—It is important to arrange the headrest so that S is

directly in front of the apparatus, with his eyes on a level with the windows in the box. The distance is less important; 50 cm. will be found convenient. The degree of illumination is controlled by two levers, which move the reflecting screens, and which are provided with scales upon the upper surface of the box. *E* first sets the right-hand lever at the point which affords the maximal illumination of the right-hand window, and records the scale-reading exactly. In accordance with the methods just outlined for the discrimination of grays, *E* now determines the just discriminable difference in the setting of the two levers (when either one of them is at the maximal point). The same precautions must, of course, be taken to reverse the standards in order to correct the space error.[1]

Variations of Method.—Substitute a 32 C. P. ruby lamp for the frosted lamp, and determine the discriminative capacity for reds of different brightness. Other colors may be employed similarly in this apparatus.

Results.—(1) *Trained observers*, working under conditions similar to those prescribed, can discriminate a brightness difference of $\frac{1}{120}$, though this fraction is appreciably altered by changes of technique or of experimental conditions. Untrained observers have less efficiency, about $\frac{1}{30}$, according to Spearman.

(2) By a different method and with colored stimuli, Gilbert found that discriminative ability increases very gradually up to the *age* of 17, but exhibits marked irregularities at the age of 7.

(3) In discrimination of shades of color, one may conclude from studies by Nichols (5), Gilbert (3), and Thompson (7), that women and girls very slightly exceed men and boys in this capacity. Luckey, however, concluded that no *sex differences* could be demonstrated in color discrimination.

(4) Individual *S*'s are apt to possess a constant *space error*, *i.e.* to tend to judge the gray on one side darker; in some cases this is the right, and in others the left, but it seems impossible to correlate this asymmetry with right and left-handedness (Spearman).

(5) Gilbert found no very decided correlation between *visual*

[1] Since the scales are identical and the entire instrument is symmetrical, a given setting of the lever will produce the same intensity of illumination for either window.

discrimination and intelligence. Spearman's experiments upon 24 village-school children give correlations between brightness discrimination and common sense, school cleverness, and general intelligence in the neighborhood of + 0.50. In a series with high-class preparatory-school boys, however, school place and brightness discrimination gave only + 0.13 for the 'raw' correlation.

NOTES.—It is imperative that the conditions under which the gray strips are observed should be kept as constant as possible. Backgrounds, cards, and holder provide these conditions in part, and relative brightness is not affected within a fairly wide range of illumination: nevertheless, it is desirable to work in the same place, at the same time of day, and under closely similar conditions of outdoor illumination, *e.g.*, between 9 a.m. and 3 p.m. on sunshiny days, and at a north window. To ensure evenness of illumination and absence of any shadows, *E* should test the setting of the experiment by placing Card No. 0 in the holder and reversing its position several times. As this card represents objective equality, any constant judgment of difference may serve to indicate uneven conditions of illumination.

In working with brightness differences, and indeed, with all small differences, *E* must be very careful to avoid suggestion of the direction of the difference to *S*, and must keep a persistent watch for all kinds of secondary criteria of judgment. If desired, one could experimentally determine the degree of objective brightness difference that could be overcome by suggestion.

REFERENCES

(1) W. Ament, Ueber das Verhältniss der ebenmerklichen zu den übermerklichen Unterschieden, etc., in Ph. S., 16: 1900, 135–196.

(2) J. Fröbes, Ein Beitrag über die sogenannten Vergleichungen übermerklicher Empfindungsunterschieden, in Z. P., 36: 1904, 344.

(3) J. A. Gilbert, (*a*) Researches on the mental and physical development of school children, in Yale S., 2: 1894, 40–100. (*b*) Researches upon school children and college students, in Iowa S., 1: 1897, 1–39.

(4) G. Luckey, Comparative observations on the indirect color range of children, adults, and adults trained in color, in A. J. P., 6: 1895, 489–504.

(5) L. Nichols, On the sensitiveness of the eye to colors of a low degree of saturation, in Amer. J. Science, 30: 1885, 37.

(6) C. Spearman, General intelligence objectively determined and measured, in A. J. P., 15: 1904, 201–293.

(7) Helen B. Thompson, The mental traits of sex, Chicago, 1903. Pp. 188.

(8) E. B. Titchener, Experimental psychology, vol ii., N. Y., 1905.

(9) E. Toulouse, N. Vaschide, and H. Piéron, Technique de psych. expérimentale, Paris, 1904, Pp. 330.

(10) J. van Biervliet, L'asymétrie sensorielle, in Bull. Acad. Royale des Sciences, etc., de Belgique, 34: Série iii., 1897, 326–366.

TEST 18

Auditory acuity.—This test, like that of visual acuity, is primarily conducted for hygienic and practical purposes, especially in the examination of the physical condition of school children, and constitutes the chief auditory test. We may distinguish between simple acuity tests, which are designed merely to detect the existence of lessened aural efficiency and roughly to measure its degree, and more elaborate tests of a diagnostic character, which are for the most part, not used in group investigations, but are confined to the work of specialists in otology or in the psychology of audition.[1] Among the latter tests may be mentioned that of binaural pitch-difference, integrity of the tonal scale, bone vs. air-conduction, determination of relative and absolute deafness, diagnostic speech-tests, etc. These tests are designed to investigate the functional efficiency of the various auditory structures, such as the tympanum, ossicles, cochlea, auditory nerve, and to determine the cause of the defect in hearing and the possibility of alleviating it by medical treatment. In particular, it is important, from this point of view, to differentiate between defect in the middle, and defect in the internal ear, because in the former case partial deafness may often be relieved, whereas in the latter medical treatment is ordinarily of no avail.

The more common and widely employed tests for acuity fall naturally into two main groups, viz: speech tests and instrumental tests. Speech tests may be conducted by either vocalized or whispered speech, and by either the method of extreme range or the method of percentage of accuracy. For instrumental tests, use is most often made of the watch, of some form of audiometer, or

[1] A typical illustration is given by the interesting article of Bingham (4.)

acoumeter, or of a tuning fork. The relative merits of these tests deserve brief consideration.

The primary advantage of *speech tests* is that they measure directly the most important function of the ear—the hearing of conversational speech, whereas all instrumental tests, because they test the perception of only a limited number of auditory qualities, fail to give unequivocal indication of auditory efficiency. One may hear the watch at a considerable distance and yet be relatively deaf for speech, or conversely. Speech tests should, accordingly, be given the preference where possible.

The use of speech tests is, however, rendered difficult for several reasons. (1) Speech involves a great variety and complex combination of pitches of varied intensity and clang-color, and these elements are further varied by changes in accent, emphasis, and inflection. To render speech tests available, therefore, most careful study must be made of the elements of spoken and whispered speech, and lists of test-words must be prepared in the light of this analysis.[1]

(2) Examiners can not guarantee uniformity of enunciation and intensity of stress in conducting the test, so that the results of different *E*'s, or even of the same *E* at different times, are rendered difficult of comparison. This difficulty must be met both by preliminary practise and care on *E*'s part, and by ranking *S*'s relatively, in terms of the empirically determined norms for each particular test.

(3)The acoustic properties of the room in which the test is held markedly affect its outcome. The method of relative ranking, coupled with the method of constant range (described below), must be used to meet this difficulty.

(4) Unavoidable noises are more likely to interfere with speech tests than with tests conducted at close range, *e. g.*, by the audiometer. To offset this, tests must be conducted in as quiet a room as possible, and doubtful cases must be retested under the most favorable conditions that can be secured.

Limits of space will usually preclude the use of vocalized or conversational speech, but *whispered speech* may be used for tests in a range of from 17 to 40 m., or about one-third that of vocalized speech. Whispering reduces the intensity of the vowels, whereas consonants are little changed. This test serves perfectly well for the practical examination of hearing and should be employed whenever feasible.

In the use of both speech and instrumental tests it has been customary to employ the *method of extreme range*. A range line is chalked off on the floor of the room; *S* is seated at one end of this range, while *E* moves methodically forward and backward over it, until he determines the extreme limit of

[1] This work has been done by Wolf (20). English number-word lists have been prepared and tested by Andrews (1). Reference to these writers will make clear why disparate words form the best speech-test material, and why numbers form the best type of words. Politzer's objection to numbers (11, p. 117) is answered by Bezold (2, p. 5; 3, p. 206).

auditory capacity for the voice or instrument. The careful experiments of Andrews (1) have revealed most serious errors in this method, due to variation in the reflection of sound waves when *E* changes his position in the room. In other words, intensity does not decrease in any uniform manner as distance increases, and consequently there is no constant relation between length of range and goodness of hearing. It is probable that this error has entered into practically all tests conducted indoors by this method. The only way to avoid it is to use the *method of constant range*, *i. e.*, to keep the distance of the range, and thus the acoustic conditions, constant, and to measure acuity in terms of the percentage of errors made in a series of tests at this selected range. Furthermore, since there are distinct differences in the audibility of different syllables, it is imperative to employ only selected lists of test-words, and to employ a sufficient number of them to include all the desired vocal elements.

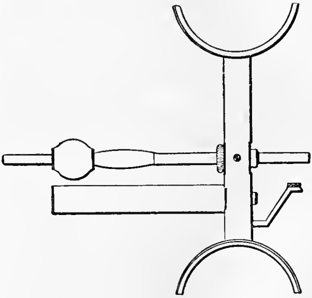


FIG. 42. POLITZER'S ACOUMETER.
From Titchener, *Experimental Psychology*.

Reverting to *instrumental tests*, we find that the *watch* is most widely used. Its advantages are its convenience and accessibility and its relatively short range. Its disadvantages are that, like any instrument, it fails adequately to test the capacity to perceive speech, that its sounds give rise to a perception of rhythm, that its ticking is so familiar that illusions of hearing are frequent, and that watches vary in the intensity and quality of their ticks.[1]

Various forms of *acoumeter* have been invented to meet the deficiencies of the watch. The instrument invented by Politzer (Fig. 42) is best known

[1] Statements sometimes made in books on hygiene that, if the ticking of a watch can be heard at so-and-so many inches, the subject has normal hearing, are obviously absurd. The normal range for a watch-tick is given at 2.5 to 4.5 m., but one in the author's possession has a range of 12 m. See Bezold (3) and Sanford (11, p. 55).

and is extensively employed in clinical work. Its range is commonly given at 15 m., but will vary one or two meters from this, as test conditions vary. This acoumeter yields a brief tone, 512 vibs., of constant intensity. Hearing is tested by the method of extreme range.

For description, see Politzer (11, pp. 107–8). The upright is held between the thumb and forefinger, and the small hammer is dropped upon the steel cylinder from a constant height. A small disk attached to a pin, not shown in the cut, is used for bone-conduction and other diagnostic tests.

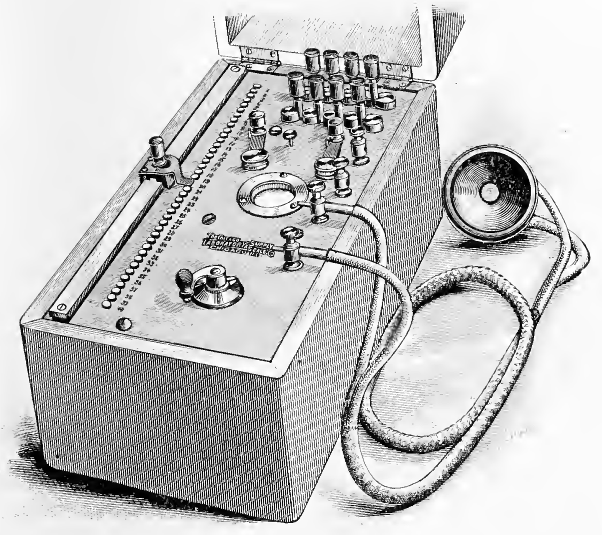


FIG. 43. SEASHORE'S AUDIOMETER.

Lehmann's acoumeter (Fig. 44), which is here prescribed, has the advantage of allowing variation of intensity, and is thus adaptable to the space limits of the ordinary laboratory.[1]

The acoumeter described by Toulouse, Vaschide, and Piéron (15) substitutes a drop of distilled water for the metallic ball, and an aluminum disk for the receiving plate.

[1] For description, see Hansen and Lehmann (7).

Many attempts have been made to devise an instrument that will permit testing at the ear itself, in order the better to rule out disturbing noises. Commonly, these devices are electrical in nature, and are planned to utilize a telephone receiver in which clicks or tones are produced in a graded series of intensities. Typical of these instruments is Seashore's *audiometer* (Fig. 43), which has been fully described by its inventor (14), and which has been extensively employed by him (15) and by others, *e. g.*, by the Child Study Bureau at Chicago (9, 16). The results that have been obtained by all instruments of the telephone type have apparently been rendered unreliable by physical errors (particularly by variations in the sensitivity of the telephone receiver), which are difficult to eliminate.[1]

Tuning forks may be employed for acuity tests in accordance with the method first suggested by Von Conta (19), in which a 512 vibs. fork is struck and brought before the ear to be tested, and acuity determined by the length of time it can be heard. Blake's fork (Fig. 45) is devised especially for use by such a method, and may also be used for simple diagnostic tests as described in detail below.

A. WHISPERED SPEECH TEST

Materials.—Meter stick. Telegraph snapper, for signalling. A number of small rubber stoppers, for ear plugs. List of 100 test-numbers arranged in ten series, as in the following Table.

TABLE 25

Test-numbers for Auditory Acuity (Andrews)

I	II	III	IV	V	VI	VII	VIII	IX	X
6	84	19	90	25	14	8	52	73	24
29	69	53	7	13	31	93	35	41	95
42	17	34	39	46	9	27	64	16	62
87	92	28	62	7	65	60	81	95	49
53	33	97	84	54	98	15	6	57	80
94	26	45	21	70	76	74	19	38	71
70	50	72	56	91	40	36	78	20	16
35	75	60	75	83	23	49	40	89	3
18	48	3	43	68	52	82	23	64	58
61	1	86	18	92	87	51	97	2	37

Preliminaries.—Select, if possible, an oblong room of average proportions and a length of at least 30 m. By rough preliminary

[1] For an extended discussion of the tecnique, and particularly of the calibration of this type of apparatus, consult Bruner (5).

tests, establish a range in this room such that not over 90 of 100 test-words can be correctly heard by a normal ear. If space will not permit this range to be established otherwise, interpose screens between *E* and *S*, or place *E* and *S* in adjoining rooms, off a straight line. The range may thus be cut down to from 18 to 20 m., or even less. Whatever may be the arrangement that affords a suitable range, make careful note of all acoustic conditions, *e.g.*, distance of range from walls, dimensions of rooms, exact position of *E* and *S*, disposition of large pieces of furniture in the rooms, number of doors or windows opened or closed, time of day, etc. Be sure always to work under precisely these conditions.

METHOD.—(1) Seat *S* at the end of the range selected, with his right ear toward *E*. Carefully close the left ear by means of a rubber stopper inserted into the meatus.

This must completely close the ear, but must not be distressingly tight. *E* should practise on himself beforehand. If both ears are properly stopped, the ticking of a fairly loud clock can be heard only with difficulty when 1 or 2 m. away, and an ordinary watch cannot be heard when held close to the ears. The plug of cotton often used is entirely inadequate. Inserting the moistened finger-tip into the meatus makes an effective plug, but the position is uncomfortable, and *S* is likely to move the finger and thus to cause distracting noises in the stopped ear. The same objection may be made to the practise of stopping the ear by pressing in the tragus, or by closing the meatus with the fleshy part of the ball of the thumb.

Direct *S* to close or shield his eyes during the test, and on no account to watch *E*'s lips. His mouth must likewise be closed, since hearing is altered when the mouth is opened.

Give *S* a short preliminary series without recording results, until satisfied that he understands the conditions of the test and feels at ease.

(2) For the more formal test, pronounce the 100 words (or but 50, if time is limited) in groups of 10, in the following manner: at the conclusion of one expiration, snap the sounder once as a ready signal for *S*: at the conclusion of the next expiration, pronounce the test-number in whispered speech with the residual air in the lungs: then snap the sounder twice to indicate that the word has been pronounced, and let *S* either speak or write down the number that he has heard (using a dash if nothing is heard). Meantime, *E* inter-

polates three complete breaths, then gives the warning signal, then the test-number after the fourth breath, and so on until 10 test-numbers are given. After a brief rest, try the second 10 numbers, and similarly, the third, fourth, etc. To avoid possible error, let *S*, if he is writing his report, begin a new column with each ten.

(3) Stop *S*'s right ear and test his left ear in the same manner.

(4) Test *S*'s binaural hearing by letting him *face E*, but with precaution that he does not secure visual aid from *E*'s lips. This test is important, because binaural hearing may not be related to monaural range, and it is the type of hearing actually used in daily life. If time is very restricted, test this form of hearing alone.

Treatment of Results.—*S*'s acuity is determined by the percentage of test-numbers correctly heard, in relation to the normal percentage which has been ascertained by averaging the percentages of all *S*'s tested under the same conditions. Thus, if the normal percentage be 70, and *S*'s be 60, his acuity is 6/7; if *S*'s be 80, his acuity is 8/7, *i.e.*, supra-normal. Credit may be allowed for partially correct reports, *e.g.*, 62 for 65: such allowance is specially recommended if 50 or fewer test-numbers are used.

Notes.—The sounder is used to avoid changing *E*'s vocal 'set.' If during the test, *S* becomes restless or inattentive, defer its completion.

It is best to test but one *S* at a time: two *S*'s may, however, be placed back to back, for testing the right ear of the one and the left of the other, if precaution is taken to ensure against communication or disturbance. If the room is large, and preliminary tests warrant the belief that acoustic conditions will be identical, more *S*'s may be tested by seating them on an arc equidistant from *E*.

A very crude group test may be carried out by placing all the children in a room at the limit of the ordinary classroom distance. Let them all close their eyes; then order them in a whisper to perform some unusual movement, such as to place the right forefinger on the palm of the left hand. Repeat with similar commands. Note any children who fail to respond, or who do so in evident imitation of others. Give these more careful tests later. Or take smaller groups of 10, similarly placed across the classroom. Provide each with a block of paper and pencil. Try a series of 10 whispered number-words, and let each write them as heard. Test carefully any who make a single error.

B. ACOUMETER TEST

Apparatus.—Lehmann acoumeter, provided with glass, copper, and cardboard receiving plates (Fig. 44). Small level. Meter stick.

Preliminaries.—Select a room with a straight range of at least 10 m. Seat *S* at one end of this range (10 m.) with his unused ear plugged, eyes and mouth closed, as in the speech test. Place the

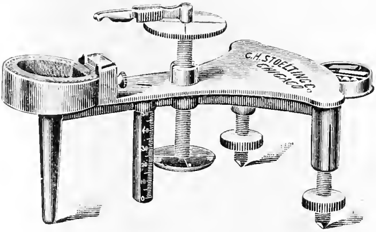


FIG. 44. LEHMANN'S ACOUMETER.
(Improved by Titchener.)

acoumeter upon a table at the other end of the range, and adjust it to a true horizontal plane by means of the test-level and levelling screw. Use the steel ball and the glass receiving plate, if only one plate can be tested.

Method.—(1) Conduct a series of preliminary trials to familiarize *S* with the test conditions. Give a verbal 'now' about 2 sec. before the ball is dropped. Let *S* report 'yes' or 'no' after each trial. Introduce a few check tests, *i.e.*, tests in which the 'now' is spoken, but the ball is not dropped. From these trials, *E* can determine approximately the 'critical' height for *S*.

E manipulates the acoumeter with his right hand, using thumb and forefinger to press the forceps, and thumb and middle-finger to turn the milled head of the screw on which the forceps rest. The shot is picked up with the left hand and placed in the rounded cavity in the tip of the forceps. Beneath the instrument will be found a vertical millimeter scale. A flat disk attached to the vertical screw just grazes this scale, while the disc is divided by cross lines into quadrants, so that variations in height of $\frac{1}{4}$ mm. or less may easily be secured. By setting the disk at zero and working upward, the height of fall may be noted without further reading of the vertical scale,

simply in terms of quarter-turns of the screw. *E* must practise this manipulation until it becomes automatic; special care must be taken to make a clean release of the shot, without swerving from the point just over the center of the receiving plate.

(2) Conduct a series of 10 trials from a constant height, so chosen as to lie probably just above *S*'s hearing capacity. If this is done, he should then report correctly all ten trials. It is understood that several check tests are added to the series of 10.

(3) Reduce the height of fall by a half turn, 0.5 mm., and give another similar series of 10 trials, with check tests added. If *S* answers correctly, reduce the height by another half turn, and continue in this manner, until there is found a set of the screw at which *S* begins to make errors. It is well to confirm the result by taking series with a still smaller fall.

Treatment of Results.—The last correctly given series may be taken as the measure of *S*'s capacity. His acuity is measured, as in the previous test, by his relation to the norm or average result determined under the same conditions. In physical terms, *S*'s capacity can be indicated by stating the conditions, distance from the instrument, material used for receiving plate, and indicating the physical measure of the noise produced, *i.e.*, in mg.-mm. The ball should be weighed upon sensitive scales to secure this index.

Typical Results.—Lehmann's results for average *S*'s at 10 m. are: with glass plate 540 mg.-mm.; with copper plate, 1110 mg.-mm.; with cardboard plate, 225 mg.-mm.[1] Other experimenters report lower limens, *e.g.*, 500 mg.-mm. with the copper plate.

Results obtained by analogous methods are those of Schafhäutl, who found that the noise made by the fall of a cork pellet weighing 1 mg. from a height of 1 mm., upon a glass plate, could be heard 91 mm., and of Nörr, who found that with small iron balls dropped upon an iron plate, the normal limen for 50 cm. distance was 1500 mg.-mm.

Notes.—All work with liminal stimuli is difficult, and this is especially true in audition. In the present test, *S* may imagine that

[1] These figures are for a 'plate' 1 x 1 cm. in size and 1 mm. thick. If a larger plate is used, as in the regular equipment, the limen is altered. Thus, with the glass plate, Lehmann and Hansen report 432 mg.-mm., with the area increased to 1 x 2 cm., and 16 mg.-mm. when the area was increased to 1 x 3 cm.

he hears the ball drop when it does not. Check tests are demanded for this reason. Occasionally an *S* may be found so 'imaginative' that the test can not be successfully used. The only remedy is to try to increase his caution by informing him of his errors.

C. TUNING FORK METHOD

APPARATUS.—Blake's fork (Fig. 45). Stop-watch. [Rubber tube.]

METHOD.—(1) Stand directly behind *S*. Sound the fork by pressing the tips of its prongs together until they touch, and then suddenly releasing them. Hold it opposite, and close to the ear to be tested, with its plane of vibration vertical. Lift the prongs

FIG. 45. BLAKE'S FORK.

For acuity and diagnostic tests by the temporal or 'ringing-off' method.

away from the ear occasionally, so that *S* can state more easily when it actually ceases to be heard. Record the time by means of the stop-watch. Repeat 5 times with each ear, or until accordant times are given. Compare this time with the norm previously established empirically for the fork in use.

(2) For a simple diagnostic test, place the stem of the sounding

fork between *S*'s teeth. If both ears are normal, *S* will hear the tone with equal intensity in each ear, or the tone may be subjectively located in the middle of the head. If, however, one ear is defective, the tone may be heard either more loudly or less loudly in the affected ear. If the tone is heard more loudly in the ear which previous tests have shown to be defective, we may expect that the location of the defect on that side is in the middle or external ear, and that it may yield to proper medical treatment. If, on the contrary, the tone is heard better in the good ear, we may expect that the defect on the other side lies in the internal ear, or in more deeply seated portions of the auditory mechanism, and that it will probably not yield to treatment.

General Results and Conclusions.—(1) It is difficult to state the prevalence of defective hearing in school children, because of the arbitrary and loose nature of the tests that have been used, and the varying standards that have been set for normality of hearing. Thus, in New York City, a recent report indicates only 1.1 per cent defective hearing; but here the test consisted merely in the use of a few whispered words in the school room at *20 feet* distance. The extensive Chicago tests,[1] conducted with Seashore's audiometer upon 6729 children, show that, if a pupil is classed as defective when the audiometer record is four points or more below the norm (indicating a defect such that "he would be seriously inconvenienced in detecting sounds of medium intensity"), 1080, or 16 per cent, of the number were defective in one or both ears (6.64 per cent in both, and 9.55 per cent in one ear). A defect equivalent to three or more points of the audiometer scale was found in one ear in 26.3 per cent, and in both ears in 12.3 per cent of those examined.

Other examinations are summarized by Young (21) as follows: "Sexton, of New York, examined 575 school children, of which 13 per cent were hard of hearing; W. von Reichard, testing with the watch 1055 pupils of the gymnasium of Riga, found 22.2 per cent with defective hearing. Weil, of Stuttgart, tested the sense of hearing in 5905 scholars of various kinds of schools, and found it below the normal in from 10 to 30 per cent of the children, according to their social condition. Moure, of Bordeaux, found 17 per cent;

[1] See Smedley (16), Macmillan (8, also summarized in 12).

Gellé, of Paris, 22 to 25 per cent; Bezold, of Munich, 25.8 per cent of pupils with hardness of hearing." See also Chrisman (6) for a summary of investigations prior to 1893.

(2) With regard to the *partially deaf*, Macmillan and Bruner (9) conclude that, in theory, there exist varying degrees of deafness, "ranging all the way from slight and temporary impairment of hearing due to a cold, to the stage of absolute and permanent silence." An examination of the children attending the public day-schools for the deaf in Chicago, however, showed a somewhat unequal division of these pupils into 5 classes, based upon the somewhat conventional and immediately practical test of the status of the pupil in hearing in his schoolwork. Thus, of 174 cases, 55 were classed as totally deaf, 33 as "practically deaf" (hearing only intense and continuous sounds), 53 as possessing "a degree of hearing power" (hearing loud sounds, but not understanding vocal speech),[1] 25 as "deaf for ordinary school conditions" (hearing only words spoken loudly and close to the ear), and 8 as "hearing children temporarily needing special training in articulation."

(3) *Differences between the two ears.* Seashore (15) found decided differences in the acuity of the two ears, differences that were unknown to the *S*'s that exhibited them. Preyer, Fechner, and Bezold have concluded that the left ear tends to be the more acute: Bruner (5), however, as well as Miss Nelson (10), state that in both sexes the right ear is the more acute. Van Biervliet (18) asserts that inequality of hearing of the two ears is a universal fact, that the disparity is such that the poorer ear has a capacity $\frac{1}{9}$ less than the better ear, but that the right ear is the better in right-handed, the left in left-handed *S*'s.

For practical purposes in connection with schoolroom tests, the determination of this difference is significant only when the inferiority of one ear is marked; in such cases, pupils should be so seated in the classroom as to bring their 'good' ear toward the teacher.

(4) Seashore's tests (15) indicate that acuity improves with *age* up to 12 years: this improvement is due partly to the development of the ear, but is slightly affected by the growth in ability to understand and to undertake the test.

[1] This class offers hope of improvement in hearing by means of mechanical devices for the intensification of speech.

(5) There are no noticeable *sex* differences, according to Seashore. Lombroso concludes that men's hearing is keener than women's.

(6) Seashore says there is "no indication that the *bright children* hear better than the dull children: there may be cases of children who are dull or are accounted dull because they do not hear well, but such cases are not common enough to be revealed clearly by our method, although there may be some indication of them." Nearly every other investigator, however, has found evidence to show that defective hearing has a positively injurious effect upon school-standing.

At Chicago (16), the examination of 5706 pupils with Seashore's audiometer showed that pupils below grade have, at every age, more cases of defect than those at and above grade, and that pupils in the school for backward and troublesome boys have a greater percentage of defect than boys of the same age in other schools. At Copenhagen, Schmiegelow found that, of 79 pupils regarded by the teachers as poorly endowed mentally, 65 per cent had defective hearing. Similarly, Gellé found 75 per cent of defect in the pupils classed as poorest. Permewan, at Liverpool, averaged the distance the watch could be heard by 203 pupils when divided into three groups—bright, average, and dull— and obtained the figures 51 inches, 47.3 inches, and 31.25 inches for these three groups, respectively. Shermunski, at St. Petersburg, by means of the whisper test, found that, among those of normal hearing, the ratio of good to poor students was 4.19 to 1; among those whose hearing was but $\frac{1}{2}$ to $\frac{1}{3}$ the normal, the ratio was 2.6 to 1; among those whose acuity was less than $\frac{1}{3}$, the ratio was 1.7 to 1.

(7) *Racial differences.* Bruner's St. Louis Exposition tests (5) indicated that the whites were clearly superior in acuity to the other races tested. The Filipinos had the poorest hearing of those tested.

(8) The simplest disturbance of hearing, if allowed to continue, may lead to serious results. In general, those who test the hearing of school children should note the condition of the ear, as well as test its capacity. Discharge of matter from the ear should be a cause for reference to medical attention.

(9) Children who are partially deaf should be guided, in their adoption of occupation, to avoid callings for which they are unfitted, *e.g.*, medicine, law, music, school-teaching, stenography, telephone or telegraph work, railroad, marine or military service.

(10) The ears of school children should be tested carefully at least once in two years.

(11) Defective hearing, like defective vision, may exist in serious degree and yet pass unnoticed by child, teacher, parents, or friends. Of the 13 per cent found defective by Sexton, only 3 per cent were themselves aware of any defect, and only one of them was known to be deaf by his teachers.

NOTES.—In testing the hearing of those who are known to be partially deaf, *e.g.*, such a group as is mentioned in (2) above, the ordinary speech or instrumental tests are not serviceable. Use may, however, be made of the telegraph snapper mentioned in the first method, or of Blake's fork in conjunction with a 'differential tube.'

The noise of the snapper can be heard by the average ear at a distance of some 150 m. or more. In testing the partially deaf *S*, it should be held slightly behind his ear, out of direct view, and employed like the Politzer acoumeter, *i.e.*, by asking *S* to give the number of 'clicks' (2 to 5) that he hears. In very young *S*'s, sufficient indication of hearing may be obtained by watching for reflex starts of the whole body, or of some part of it.

The differential tube, as used by Macmillan and Bruner (9) consists of a tube of soft rubber 100 cm. long, and 4 mm. internal diameter, fitted with hard rubber tips for insertion, one into *S*'s, and one into *E*'s ear. After *S* has been familiarized with the sound of the fork by hearing it with the base applied to his front teeth, his ears are tested one at a time by placing the stem of the sounding fork upon the tube. On account, presumably, of the longer duration of the sound, this device may be used to detect a grade of hearing even lower than that detected by the snapper.

REFERENCES

(1) B. R. Andrews, Auditory tests, in A. J. P., 15: 1904, 14-56, and 16: 1905, 302-326.

(2) F. Bezold, Schuluntersuchungen über d. kindliche Gehörorgan, Wiesbaden, 1885.

(3) F. Bezold, Funktionelle Prüfung des menschlichen Gehörorgans, 1897.

(4) W. Bingham, The rôle of the tympanic mechanism in audition, in P. R., 14: 1907, 229-243.

(5) F. G. Bruner, The hearing of primitive peoples: an experimental study of the auditory acuity and the upper limit of hearing of whites, Indians, Filipinos, Ainu and African pigmies, N. Y., 1908. (Reprinted from the Archives of Psych., No. 11). See especially 55-108.

(6) O. Chrisman, The hearing of children, in Pd. S., 2: 1893, 397-441.

(7) F. Hansen and A. Lehmann, Ueber unwillkürliches Flüstern. Eine kritische u. exp. Untersuchung der sogenannten Gedankenübertragung, in Ph. S., 11: 1895, 471-530, especially 494 ff.

(8) D. Macmillan, Some results of hearing-tests of Chicago school children, in Medicine, April, 1902.

(9) D. Macmillan (and F. G. Bruner), A special report of the Dept. of Child-study and Pedagogic Investigation on children attending the public day-school for the deaf in Chicago, Chicago, 1908. Pp. 88.

(10) Mabel L. Nelson, The difference between men and women in the recognition of color and the perception of sound, in P. R., 12: 1905, 271-286, especially 280 ff.

(11) A. Politzer, Lehrbuch d. Ohrenheilkunde, Stuttgart, 1893.

(12) Rept. com. on statistics of defective sight and hearing of public school children, in U. S., 1902, ii., 2143-2155.

(13) E. C. Sanford, A course in experimental psychology, Boston, 1895 and 1898. Pp. 449.

(14) C. E. Seashore, An audiometer, in Iowa S., 2: 1899, 158-163.

(15) C. E. Seashore, Hearing-ability and discriminative sensibility for pitch, in Iowa S., 2: 1899, 55-64.

(16) F. W. Smedley, Rept. dept. child-study and pedagogic investigation, in 46th An. Rept. Brd. Educ. Chicago, 1899-1900. Also in U. S., 1902, i. 1095-1115.

(17) E. Toulouse, N. Vaschide, and H. Piéron, Technique de psych. expérimentale, Paris, 1904. Pp. 330.

(18) J. van Biervliet, L'asymétrie sensorielle, in Bull. Acad. Royale des Sciences, etc. de Belgique, 34: Série 3, 1897, 326-366.

(19) Von Conta, Ein neuer Hörmesser, in Archiv f. Ohrenheilkunde, 1: 1864, 107-111.

(20) O. Wolf, Sprache und Ohr, Braunschweig, 1871. Also various articles in otological journals, as Arch. f. Augen u. Ohrenheilkunde, 3: Abth. 2, 35, and Abth. 1, 125; and Zeits. f. Ohrenheilkunde, vol. 20.

(21) A. G. Young, School hygiene, in 7th An. Rept. State Brd. Health of Maine, Augusta, 1892. Pp. 399.

TEST 19

Discrimination of pitch.—Like other forms of sensory discrimination, this has been employed to discover the relation between such sensitivity and general intelligence. It has sometimes been employed to estimate musical ability, and it has, of course, general psychological interest. With adults and with children over eight or nine years of age, the test is relatively easy to administer.

The available instruments are air-blown reeds or bottles, vibrating strings, as in the sonometer, and tuning forks.

Experience shows that a set of finely tuned reeds may be employed only when they have 'settled' to their permanent pitch, when they are blown by an absolutely constant source of air-supply, and when their tone-color is uniform. This renders the reed-box, such as the Appunn tonometer, out of the question, save for well supplied laboratories. Gilbert's tone-tester (1), which is constructed from an adjustable reed pitch-pipe, varies as much as five vibs. in pitch with variation in the force with which it is blown. Stern's blown-bottles or tone-variators (6) necessitate a constant air-supply, and even then do not yield pitches which correspond to the attached scales. The sonometer or monochord, employed by Wissler (12) and Spearman (5), is rather unwieldy, not always constant in pitch and tone-color, and complicated by certain mechanical difficulties, while its pitches must be computed at each test in order to guarantee correctness of the assigned vibration-rate values. The instrument is defended, however, by Spearman (5, 243f). Wissler's method of using the monochord, in accordance with which *S* was obliged to manipulate the instrument, is indefensible, and his results are worthless, as far as pitch discrimination is concerned. The use of tuning forks in which the pitch of the comparison fork is varied by weights or riders (for illustration, see Titchener, 9, i., 68) also necessitates the computation of the pitch differences by counting beats, and both this and the manipulation of the riders is not easy for inexperienced *E*'s. For these reasons, a series of carefully tuned forks, selected for uniformity of tone color, one for each pitch desired, is here recommended, after the example of Seashore (3) in his examination of the pitch-discrimination of children. The present apparatus (Fig. 46) has been described fully by the author in conjunction with Titchener (10).[1]

APPARATUS.—Set of 11 forks—one standard fork of $426\frac{2}{3}$ vibs., and 10 comparison forks, whose rates are 0.5, 1, 2, 3, 5, 8, 12, 17, 23, and 30 vibs. below the standard. A resonance box on which the forks may be mounted as they are used. (Fig. 46). Soft-tipped hammer for striking the forks.

METHOD.—(*a*) *Preliminary trials.* Seat *S* with his back to the table at which *E* works, and about 1 m. distant. Instruct him as follows: "When I say 'now,' close your eyes and listen carefully to

[1] Since the apparatus here prescribed was completed, Seashore has proposed to substitute for the resonance box, specially tuned resonators (perhaps two or three in number) of the Helmholtz type. By strongly re-enforcing the fundamental tone, these resonators might be particularly valuable in eliminating chance differences in tone color which are likely to appear in the small forks in use here. Otherwise, these differences must be eliminated by careful selection of the forks.

the two tones you will hear; then tell me whether the second tone is higher or lower than the first. Say 'higher' if the second tone seems pitched above the first, 'lower' if below."

S's judgment must always be in terms of the second tone. To request him to answer merely "same" or "different," as some investigators, *e.g.*, Gilbert, have done, would produce different results, as, in general, it is less difficult to judge a difference than to judge the direction of this difference.

In this preliminary series, *S* may be allowed to give the answer "same," if he naturally does so when he is unable to say "higher" or "lower."

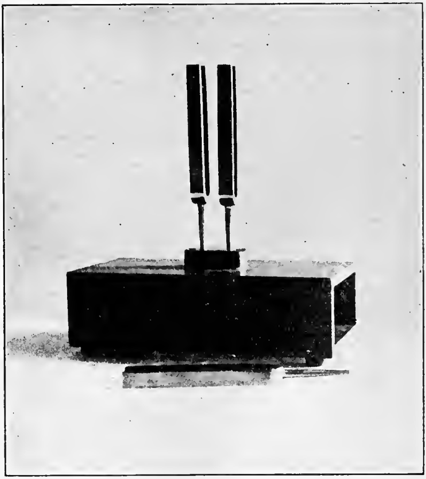


FIG. 46. TUNING FORKS FOR PITCH DISCRIMINATION.

S's who are extremely unfamiliar with tones occasionally do not understand what is meant by 'higher' and 'lower,' and, like markedly unmusical or tonally-deaf *S*'s, are apt to search for differences in intensity or duration of the tones instead of for qualitative (pitch) differences. In such an event, *E* must select two forks that give the maximal difference, and give *S* a short course of training by striking the forks in succession, and explaining after each pair that the second was higher or lower, as the case may be. If this training is futile, *S*'s discrimination must be extremely poor, and he may be ranked 30+.

Insert firmly in the oak pedestal, with their axes at right angles to the main axis of the resonance box, two forks that afford a large stimulus difference, *e.g.*, the standard and the lowest fork (marked 30). Damp one fork, *e.g.*, the nearer one, by placing the left forefinger on the tip of one of its prongs: sound the second fork by striking one prong a clean tap of moderate strength at a point about ¼ the distance from its tip. Let the fork ring 4 sec., then damp it by resting the middle finger upon it. After an interval of 2 sec., lift the forefinger, sound the first fork (while the 2d is still damped) and damp it similarly at the end of 4 sec. Keep these time relations—4 sec. 1st tone, 2 sec. interval, 4 sec. 2d tone—constant, and strike the forks as uniformly as possible in all tests. *S* always judges in terms of the second tone.

Continue the practise series, in accordance with the general plan for discrimination work described in the opening pages of this chapter, by inserting other comparison forks in place of the "30" fork, giving sometimes the standard, sometimes the comparison stimulus first. This preliminary series is to familiarize *S* with the general nature of the test, and to afford *E* a rough notion of the limits of *S*'s discriminative capacity. For most *S*'s, at least 4 or 5 min. should be given to this practise.

(*b*) *Test proper.* If *S*'s 'critical region' is not yet evident, give a formal series of pairs of stimuli, beginning with a supraliminal difference and passing toward subjective equality, until a difference is reached which *S* mistakes, or recognizes with difficulty. Keep this pair of forks on the resonance box, and give a series of ten pairs of stimuli—five with the standard first, five with the comparison first, but in an irregular order. In this series *S* must not be allowed to judge "same," but should be made to guess in case of doubt. He should know that two different forks are being used, but should not know the direction of the differences which he is judging. If he gives 10 correct judgments, select for the next series a comparison fork nearer the standard in pitch; if he gives but 5 or 6 correct answers, and these with difficulty, select a comparison fork farther from the standard. Seek a pair of forks which will yield about 8 right cases in 10. Confirm the difference limen thus secured by trying series with comparison forks just sharper and just flatter than the one in hand. *S*'s discriminative capacity for pitch is indi-

cated by the difference in vibrations between the standard fork and the comparison fork that yielded 80 per cent right answers.

RESULTS AND CONCLUSIONS.—(1) The *difference limen* for highly practised S's in careful laboratory tests is, for this region of the tonal continuum, about 0.3 vib.[1]

(2) There are large *individual differences* in the difference limen for pitch: some, even young children, can discriminate 2 vibs. with certainty; some are virtually tone-deaf. These differences are to be referred, in the main, to structural differences in the sense-organ. Some idea of the distribution of capacity in this test is afforded by Seashore's results, Table 26.[2]

TABLE 26

Pitch Discrimination of 167 Children, Aged 6–15 Years. (Seashore.)

NO.	LIMEN IN VIBS.	NO.	LIMEN IN VIBS.
20	1 to 2	21	12 to 30
63	3 " 5	14	Over 30
48	6 " 10		

(3) Seashore found no obvious *sex differences* in the discrimination of pitch. Both Wissler (12) and Thompson (8) found women superior to men, but their results are less reliable than those of Seashore.

(4) The relation of pitch discrimination to *age* is not entirely clear. If we accept Gilbert's results, they indicate a correlation between age and capacity to the extent that children are least efficient at 6, improve rapidly up to 9, then gradually to 19, with exceptions at the ages of 10 and 15—a seeming loss of capacity, which Gilbert is inclined to refer to pubertal and other physiological disturbances.[3] Wissler's seniors surpass themselves as freshmen.

[1] For a summary of the work of Delezenne, Seebeck, Preyer, Luft, Meyer, and others, consult Titchener (9, Pt. ii., 235 ff.)

[2] These results are subject to the qualification that the figures obtained from those of the children under 9 years of age are not very reliable. In some 30 cases not here included, Seashore was unable to determine a limen within the time at his disposal. Gilbert found only 3 of 130 children who could not discriminate a half-tone, *i.e.*, about 30 vibs. in this region, but as already stated, his S's were asked to judge only difference.

[3] Cf. similar disturbances found by Bryan at 10 and 15 in motor tests.

Seashore, however, discards results obtained from children under 10, and can find no certain indication of improvement in discrimination after that age.

(5) Seashore found a slight positive correlation between pitch discrimination and *auditory acuity* (Test 18).

(6) (*a*) *Practise* does undoubtedly improve pitch discrimination, but investigators are not in agreement as to the extent of such improvement. It seems evident that its limits are fixed by anatomical and physiological conditions in the ear itself, and that these limiting conditions vary in different individuals. In general, the improvement is not as great as that observed in some other functional capacities, *e.g.*, the discrimination of dual cutaneous impressions (Test 23), and is reached after a relatively short period of training.

Seashore believes that maximal capacity can be attained after very little practise. In 20 days training, he found that some *S*'s exhibited no improvement, while the maximal improvement reported was the reduction of the limen of an unmusical *S* from 30 to 5 vibs.

Spearman (5, p. 231) believes that 15 minutes fore-exercise will reduce the pitch-limen by an amount depending very largely upon *S*'s previous general familiarity with tonal experiences. Thus, he computes a reduction for specially practised *S*'s from a limen of 0.5 (before the 'exercise') to 0.3 (after the special exercise), for musicians from 4 to 2 vibs., for non-musicians of general culture from 10 to 4 vibs., and for European villagers from 30 to 8 vibs. These figures would indicate that even a practised musical *S* profits by a preliminary 'warming-up,' and they emphasize the importance of giving such fore-exercise to all *S*'s.

(*b*) Aside from practise in the narrow sense, *i.e.*, special exercise in pitch-discrimination under experimental conditions, we may consider the effect upon discrimination of practise in the wider sense, *i.e.*, of general *musical training*. Seashore is, again, very emphatic in his declaration that individual differences in pitch discrimination are not due principally to musical training, and Spearman's conclusion is that, "though a correspondence really does exist, yet it is not to the smallest degree of the specific character contemplated by those who talk of 'musical sensitivity,' " *i.e.*, by those who refer to pitch discrimination as a test of "musical sensitivity."

To the author, this seems a case of one-way correspondence: an individual who cannot discriminate a half-tone cannot be musical, but an individual

who is not musical *may* have a perfectly good discrimination after a little preliminary practise. Given, however, a good natural capacity for discrimination, it is unquestionably true that musical training tends to keep this capacity up to the individual's physiological limit. In testing 50 grammar-school boys for pitch discrimination, the three best discriminators were found to be "taking lessons" on the violin. The author has also shown elsewhere (11) that in the case of an unpractised, unmusical *S*, it may be possible to reduce the limen very decidedly by working under very constant, favorable conditions—such as duration, intensity, timber of tones, time-intervals, etc., but that the slightest modification of these conditions will make discrimination very difficult or impossible.

That one can argue from capacity in pitch discrimination to capacity to profit by musical instruction is asserted specifically by Seashore in another article (4, 76–7), where he says that, if the limen for pitch is .04 tone (about 2 vibs. in the section of the tonal continuum under test), the child may become a musician; if .06 to .16 tone, the child should have a simple musical education, including obligatory singing in the school; if .18 to .34 tone, the child should have this musical education only provided special inclination for some kind of music is displayed, while participation in school singing should be optional; if the limen is .36 tone (18 vibs.) or over, the child should have nothing to do with music.

(7) Seashore reported no correlation between *pitch discrimination and intelligence*, when general intelligence was indicated by class standing and teachers' estimates, and the correlation was worked out by the method of group-classification. Spearman terms this an "ingenious," but somewhat "disseminated" method, and by subjecting Seashore's results to his (Spearman's) methods, obtains from them a correlation index of 0.24 ± 0.07. From his own results, Spearman (5) concludes that general intelligence is correlated with pitch discrimination by the index 0.94, or, as he states, "The [Intellectual] Function is 9 parts out of 10 responsible for success in such a simple act as Discrimination of Pitch."[1] Later, in conjunction with Krueger (2), Spearman computes a correlation of .83 between the capacity for pitch discrimination and the hypothetical "central factor," which is tentatively assumed to be a certain quality or degree of plasticity of the central nervous system.

The author's tests of mental and physical ability in fifty 8th-grade boys included a determination of pitch discrimination, both with the Stern tone-variator and with a monochord. These two tests

[1] It is but fair to call attention to the fact that Spearman's formulas have been called in question.

showed a correlation of .83. The variator test, which was, on the whole, most reliable, showed a correlation of .27 with class standing.

(8) *Other correlations.* Krueger and Spearman report 'raw' correlations between pitch discrimination and both adding and the Ebbinghaus completion test, of .67 and .59, respectively. After the application of the expanding formula (No. 41, Ch. 3), these correlations become .80 and .81, respectively. Thus it appears certain to these authors (2, p.78) that the capacity to discriminate pitch actually exhibits a very high degree of correlation with the seemingly fundamentally different capacities requisite in adding and in the Ebbinghaus test.

The author found a correlation of .27 between pitch discrimination and the discrimination of lifted weights.

NOTES.—With reference to musical ability, Stumpf (7, ii., p. 157) proposes as tests of musical capacity: (1) discrimination, (2) ability to sing a note struck on the piano, (3) ability to judge whether one or two tones are present in various fusions, (4) skill in determining the relative pleasantness or unpleasantness of two chords separated by a short pause. M. Meyer denies the validity of the discrimination test for musical ability, and favors a form of test in which *S* is asked to state whether a given bass note does, or does not, form the proper fundamental for a given chord (played in the treble region of the piano).

It is well to inquire of all *S*'s, before the test is administered, whether they are musical or not, whether they play any musical instrument, or sing, or are 'fond' of music. The author has found several instances of children who were quite unable to distinguish pitches several tones apart, but who were compelled by their teachers to take systematic instruction in singing along with other children in the public-school classes.

S's that fail to discriminate the 30 vibs. difference may be further tested by a piano to see whether they are absolutely tone-deaf. If time permits, it is of interest to see whether *S*'s with very poor discrimination can be improved by systematic practise.

For best results, the discrimination test should be given individually; if necessary to undertake group tests, it is better to work with small groups of 5 or 6 *S*'s: supply them with pencil and paper; let them number the trials and write their judgments—"H" for higher or "L" for lower—after each number.

REFERENCES

(1) J. A. Gilbert, Experiments on the musical sensitiveness of school children, in Yale S., 1: 1893, 80–87.

(2) F. Krueger and C. Spearman, Die Korrelation zwischen verschiedenen geistigen Leistungsfähigkeiten, in Z. P., 44: 1907, 50–114.

(3) C. E. Seashore, Hearing-ability and discriminative sensibility for pitch, in Iowa S., 2: 1899, 55–64.

(4) C. E. Seashore, Suggestions for tests on school children, in Educ. Rev., 22: 1901, 59–82.

(5) C. Spearman, General intelligence objectively determined and measured, in A. J. P., 15: 1904, 201–293.

(6) L. Stern, Der Tonvariator, in Z. P., 30: 1902, 422–432.

(7) C. Stumpf, Tonpsychologie, 2 vols., Leipzig, 1883, 1890.

(8) Helen Thompson, The mental traits of sex, Chicago, 1903. Pp. 188.

(9) E. B. Titchener, Experimental psychology, vol. ii., Quantative experiments, N. Y., 1905.

(10) E. B. Titchener and G. M. Whipple, Tuning forks for pitch discrimination, in A. J. P., 20: 1909, 279–281.

(11) G. M. Whipple, Studies in pitch discrimination, in A. J. P., 14: 1903, 553–573.

(12) C. Wissler, The correlation of mental and physical tests, in P. R. M. S., 3: No. 6, 1901. Pp. 62.

TEST 20

Discrimination of lifted weights.—We may compare two weights either by attending passively to the pressures set up when they are laid upon the skin, or by actively lifting or 'hefting' them. In the first instance, we see illustrated the procedure employed in testing discrimination of pressure (Test 21); in the second, that employed in testing discrimination of weight (in the narrow sense). The latter form of discrimination is always the keener, since to cutaneous pressure, there are added sensations contributed from muscle, joint, and tendon, particularly from the tendon. Because it is the movement of lifting the weight that aids us in estimating its comparative amount, the determination of this form of discriminative capacity is sometimes loosely termed the measurement of the 'muscle sense,' or the 'muscle sense test.'

Of the historical development of this important experiment, this

is not the place to speak.[1] Attention is given here merely to the use of the experiment as a comparative test of mental or psychophysical efficiency. The most important investigations of this type are those of Gilbert (3), Thompson (7), and Spearman (6). The test was also included in those administered by the author to 50 8th-grade boys.

Gilbert used weights of the 'cartridge' pattern, similar to those here prescribed. His method was less exact than could have been desired: his school children had simply to sort out all the weights that were the same as the standard, 82 g. As there were but 9 comparison weights, yielding a maximal range of 100 g., Gilbert encountered numerous cases (see Table 27) in which no discrimination could be made within this range.

Miss Thompson employed cartridge weights giving a range from 80 to 100 g., and apparently found no difficulty in testing adults with this equipment.

Spearman, similarly, employed cartridge weights with a standard of 1000 grains and with a series of geometrical increments, as proposed by Galton (2, Appendix). It is to be noted that the smallest increment of the original Galton series, 1/100, proved too coarse to test the capacity of some of Spearman's *S*'s, while the largest increment (mentioned by Galton for use with "morbid" cases) proved too fine to test the capacity of others.

A test carried on at Columbia University under the name "perception of weight" or "force of movement" consisted in lifting the handle of a spring dynamometer until it touched a stop. The 'reagent' then made 10 successive attempts to pull the handle to the same point when the stop was removed. It is evident that the results cited by Wissler (11) for this test are not comparable with those obtained by the standard form of the weight-discrimination test.

Van Biervliet (9) used weights of 500, 1000, 1500, and 2000 g. on the 'favored' side of the body in the case of 100 *S*'s that were tested by him for asymmetry in weight. The weights were lifted by a string attached to the index finger, and a simple gradation method was employed to determine the equivalent, for the left side, of a given weight on the right side of the body. The method is too unlike the standard method to admit of comparison of re-

[1] The stock laboratory experiment is described and its technical aspecst are discussed with sufficiently full citations of its literature by Titchener (8, Pt. I, 115 ff.; Pt. II, 265 ff.). As Titchener remarks: "This may be regarded as the classical experiment of quantitative psychology. On the psychphysical side, it has engaged a long line of investigators: Weber himself (10), Fechner (1) and Hering, all employed it to test the validity of Weber's Law; and a glance at the current magazines will show that the work begun by them has continued down to the present day. On the psychological side, it has been made by L. J. Martin and G. E. Müller (4) the vehicle of a qualitative analysis of the sensory judgment, the most elaborate and penetrating that we have."

sults. It may be stated, however, that the fraction 1/9, which the same author claims to have established as a constant of asymmetry in all sense-departments, was also found in this test, *e.g.*, a weight of 450 g. in the left was equal to a weight of 500 g. in the right hand, etc.

APPARATUS.—Set of discrimination weights, comprising a standard, 80 g., and 23 comparison weights, yielding the series—80.5, 81, 81.5, 82, 82.5, 83, 83.5, 84, 84.5, 85, 86, 87, 88, 89, 90, 92, 94, 96, 98, 100, 105, 110, and 120 g. The weights are of identical size, shape, and color; are made of wood to avoid disturbing temperature sensations, and are marked inconspicuously (with reversed numbers), so that their weight may be known to *E*, but not to *S*.

METHOD.—Follow the general plan of procedure outlined in the introductory pages of this chapter. This plan embodies (1) a preliminary series of trials between the standard and various comparison weights to familiarize *S* with the conditions of the test and to indicate to *E* the probable 'critical region' in which *S*'s limit of capacity will be found, (2) a more formal determination of this region by systematic procedure from too great to too small a stimulus-difference, (3) the selection from this region of a stimulus-difference (the standard and some single comparison weight) which may be expected to yield about 8 right judgments in 10, and which is given 10 times (5 times with the standard first, 5 with the comparison weight first—the arrangement being determined by chance), and (4) the final determination of the difference that yields 8 correct judgments in 10 by the trial of slightly smaller or slightly larger stimulus-differences, as may be required.

In the application of this procedure to weight discrimination, the following suggestions may be made. *S* should take his position, standing, before the table upon which *E* has arranged the weights. *S*'s view of the weights must be cut off, either by a well-arranged blind-fold or by a horizontal cardboard screen so adjusted that he may lift the weights easily, but may not see them. In each trial, *E* selects a comparison weight, determines upon the order (standard first or comparison first); then, with a warning 'now,' places the first weight between *S*'s thumb and his first and second fingers: *S* hefts this weight, replaces it upon the table, when *E* quickly removes it and substitutes the second weight of the pair under trial, which is, in turn, hefted and replaced by *S*. The judgment must

then be given promptly by *S* and always in terms of the second weight:—"heavier," "lighter," or "equal."[1] The details of the manner of lifting the weights may, in general, be left to each *S*. Fechner allowed 1 sec. for raising, 1 sec. for lowering, and 1 sec. for changing the weight, so that each comparison required 5 sec. for its execution.

TREATMENT OF DATA.—The difference between the comparison weight that yields 8 right judgments in 10 and the standard weight, 80 g., affords the absolute difference limen. The fraction formed by taking this difference as the numerator and the standard weight as the denominator affords the relative difference limen, and is the common index of efficiency in the test, since relative capacity is found to be constant for a given individual within a wide range of absolute weights. It is, of course, the constancy of this fraction that constitutes the essential fact of Weber's Law.

RESULTS.—(1) *Normal capacity.* In one place, Weber cites as the average sensible discrimination for lifted weights for four *S*'s, 3/32; in another place, he gives 1/40 as the difference just distinguishable by "quite the majority of human beings without any long preliminary practise" (10). Other authorities have placed the norm of performance at 1/17 or at 1/24 (Seashore, 5, p. 96). The author's tests with the apparatus and methods here described indicate for 8th-grade boys an average limen of 4.7 g. (standard 80 g.). The corresponding fraction, 1/17 , is presumably close to the average performance for boys of this age. This result is corroborated by Spearman's series, which most nearly resemble the author's in method and apparatus; Spearman quotes 1/15 for a test made under unfavorable conditions and 1/20 for a test made upon older children under favorable conditions.

(2) *Individual differences.* The work of every investigator has shown that the capacity to discriminate lifted weights differs very considerably among normal *S*'s, even when age, sex, and practise factors are eliminated. The author found 7 boys in 50 who could discriminate 80 and 81 grams, and one boy who could just discrim-

[1] In the final test of 10 trials with a constant stimulus-difference, it is preferable to ask *S* to guess in case of an equal judgment; otherwise equal judgments may be recorded as wrong. *S* should not be permitted to return to the first weight after the second has been lifted.

inate 80 and 97 grams: reference has already been made to the fact that several experimenters have found their weights inadequate to measure the wide differences in capacity that they encountered.

(3) *Dependence on age.* Spearman's tests convinced him "that the younger children were almost equal to the older ones and both were not far from adults," and also that there is no appreciable loss in weight discrimination with the coming of old age. Gilbert, however, as Table 27 shows, found a gradual improvement in discrimination from the 6th to the 13th year. Developmental disturbances appear from 12 to 14, and discrimination apparently does not improve thereafter.

TABLE 27

Dependence of the Discrimination of Lifted Weights on Age (Gilbert)

	6	7	8	9	10	11	12	13	14	15	16	17
Median Limen, Boys..........	13.0	13.2	12.2	10.2	8.6	10.2	7.6	6.0	5.0	6.2	6.0	6.0
Median Limen Girls..........	16.8	13.2	11.0	10.0	9.2	7.6	7.6	5.6	7.2	7.2	6.8	6.4
Per cent over 18 g., Boys....	26	36	35	23	12	5	0	5	0	0	2	0
Per cent over 18 g., Girls....	49	40	28	17	12	6	6	0	0	0	2	2

(4) *Dependence on sex.* Inspection of Gilbert's table shows in many, though not in all groups, evidence of superior ability on the part of boys. Miss Thompson, similarly, found men much superior to women, and is inclined to see in this relation the effect of the same factors that had made men superior in her several tests of motor ability. Spearman, however, contends that "the fluctuating differences of sensory discrimination observable in connection with sex at the various stages of growth are chiefly and perhaps altogether a mere consequence of similarly fluctuating differences in intelligence" (6, p. 261).

(5) *Dependence on practise.* The consensus of opinion is that, at least in comparison with many other mental activities, the discrimination of lifted weights is but little affected by practise. Thus, for instance, Biedermann and Loewit (quoted by Spearman)

found that a difference limen of 1/21 fell only to 1/23 at the conclusion of a protracted research. It is also true that Fechner, who devoted most heroic amounts of time to weight discrimination, did not thereby attain remarkable capacity. On the other hand, Spearman believes that nearly every *S* exhibits practise-improvement in the first stages of a test (say during the first 15 min.), and that in some cases "the improvement is enormous."

(6) *Correlation with intelligence.* Spearman's 'corrected' index of correlation between weight discrimination and general intelligence is 0.44. Gilbert apparently determined no correlation between these two factors. The author's tests show no correlation with class standing.

(7) *Other correlations.* The peculiar test of "perception of weight" used at Columbia University is shown by Wissler (11) to exhibit no correlation with accuracy of movement in striking dots or with accuracy in the 'size test' (drawing a line equal to a 5 cm. directly observed standard).

(8) A '*constant error*' is exhibited by most *S*'s, in that they tend to overestimate the second weight. This, of course, makes it doubly imperative that the procedure be so arranged as to reverse the time-order in half the trials.[1]

REFERENCES

(1) G. T. Fechner, Elemente der Psychophysik, esply vol. I. (Reprint, Leipzig, 1889, 1907.)

(2) F. Galton, Inquiries into human faculty, New York, and London, 1883. (Reprinted in Everyman's Library, No. 263.)

(3) J. A. Gilbert, Researches on the mental and physical development of school children, in Yale S., 2: 1894, 40-100.

(4) Lillie J. Martin and G. E. Müller, Zur Analyse der Unterschiedsempfindlichkeit, Leipzig, 1899.

(5) C. E. Seashore, Elementary experiments in psychology, N. Y., 1908. Pp. 218. See especially Ch. viii.

(6) C. Spearman, General intelligence objectively determined and measured, in A. J. P., 15: 1904, 201–293.

(7) Helen B. Thompson, The mental traits of sex, Chicago, 1903.

[1] On other qualitative aspects of this test, see especially Martin and Müller (4.)

(8) E. B. Titchener, Experimental psychology. Vol. II. Quantitative experiments, N. Y., 1905.

(9) J. J. van Biervliet, L'Asymétrie sensorielle, in Bull. Acad. Royale des Sciences, etc., de Belgique, 34: Série iii., 1897, 326-366, esp. 330.

(10) E. H. Weber, Der Tastsinn und das Gemeingefühl, in Wagner's Handwörterbuch d. Physiol., vol. 35, 1846. (Reprinted Separately 1869, 1851.)

(11) C. Wissler, The correlation of mental and physical tests, in P. R. M. S., 3: No. 6, 1901. Pp. 62.

TEST 21

Discrimination of pressure.—The determination of the difference limen for pressure, like that for lifted weights (Test 20), has constituted one of the standard psychophysical experiments since the time of E. H. Weber, who utilized it in connection with other tests to establish the well-known law that bears his name. By experimenting with standard weights of 32 oz. and 32 dr., respectively, Weber was able to report that "a difference of the smaller weights is not less accurately distinguished by touch than the same difference of the larger weights."

This test with 'resting weights,' sometimes termed the 'pressure sense' test, appears to have been less frequently used for functional and comparative purposes than the test with 'lifted weights.' Its feasibility depends very largely upon the type of apparatus employed. Differences between the temperature of the weights and that of the skin, variation in the temperature of the weights themselves, in the 'jar' of application, in the area and place of application, etc., must be excluded, since they inevitably produce conflicting results. To obviate these errors and to render the test more simple in execution and more reliable in outcome, the use of a 'pressure-balance,' following the principle adopted by Merkel (7, p. 255) is desirable, if not essential.

Other forms of pressure-balance have been elaborated by Jastrow (5)—figured by Sanford (8, pp. 417–8) and by Titchener (10, pt. ii.)—and by Bolton and Withey (1). The balance here prescribed has been designed by the author (11) to supply in a single relatively simple apparatus a device for determining both the capacity for pressure discrimination and sensitivity to pain (Test 22). It may be regarded as a combination of the principle of Merkel's and of Jastrow's pressure-balances and that of Gilbert's balance-algometer (3).

Apparatus.—The author's pressure-pain balance (Fig. 47). Cardboard screen with suitable supports. Seconds' pendulum

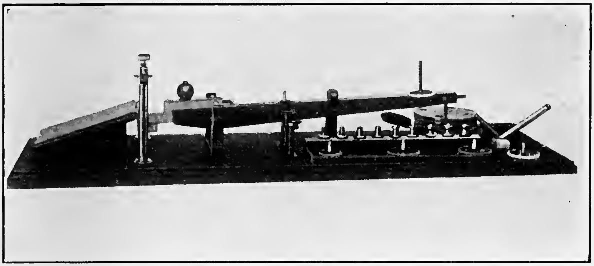


Fig. 47. pressure-pain balance.

(Fig. 21), or other noiseless device for controlling the time-relations of the test.

Preliminaries.—Place the balance upon a low table. See that the beam of the instrument moves freely, but comes to rest in a horizontal position when no weights are applied: if necessary, turn the small screw in the tip of the arm inward or outward until this position of rest is secured.

To assure comfort, the instrument should be so placed that *S*'s wrist will come just over the edge of the table; his elbow will not then be forced up into an awkward position, and his hand can lie upon the hand-rest, with the end of his forefinger projecting straight forward between the upper (stationary) and the lower (movable) tip of the balance. Adjust the upper tip so that it is in permanent contact with the center of the finger-nail, but does not touch the skin of the finger.

Arrange the screen to cut off *S*'s view of the apparatus, or, if he be reliable, simply instruct him to close his eyes.

Place the pendulum where its oscillations will be easily visible.

Method.—The general plan of procedure is identical with that outlined in the introductory pages of this chapter, and recapitulated in Test 20. To apply this procedure to the test with the pressure-balance, after throwing the lever down to the right, place

the weight marked *B-100 g.* on the pin marked *B*, at the outer end of the beam. This weight is not removed during the experiment, and constitutes the standard stimulus. Place upon the second pin, marked *A*, the desired increment weight—any one, or any combination, of the weights marked *A*. To apply a pressure-stimulus, move the release-lever up to the left, so as to depress the support beneath the beam of the balance. To remove the stimulus, move the same lever to the right. The increment-weights are added to the standard stimulus when they rest upon the beam at *A*: they are subtracted from the total pressure, at will, by depressing the increment-weight lever, which lifts them from the beam and allows only the standard stimulus, 100 g. to be operative. Thus, for example, to test the discrimination of 150 g. and 100 g., move the release-lever down to the right, place upon the pin *A* the 30 g. and the 20 g. weights, and upon the pin *B* the 100 g. weight. Give *S* a warning "now," and 2 sec. later move the release-lever smoothly up to the left: allow the pressure (150 g.) to be felt for 2 sec., then move the release-lever to the right: immediately depress the increment-weight lever, and apply the second stimulus (100 g.) in the same manner, while this lever is held down. *S* judges, always in terms of the second stimulus, saying "heavier," "lighter," or "equal."[1]

The exact duration of the stimuli and of the interval between them is of less importance than constancy from trial to trial.

To avoid local fatigue, at least 15 sec. should elapse between successive judgments.

E must practise the manipulation of the instrument, and take particular precaution to move the release lever so as to avoid either too sudden application, which produces a disturbing 'bump' and vibration, or too slow application, which also renders the judgment more difficult.[2]

S must be specially instructed to receive the stimulus passively, so far as his finger is concerned. A downward movement of

[1] Equal judgments, as previously explained, are to be avoided, if possible, in the final trials with a constant stimulus-difference.

[2] The author has found that some *E*'s, by a curious kind of unconscious 'sympathy,' are inclined to apply the light pressure more gently than the heavier pressure. *S*'s judgments will almost certainly, even without his knowledge of it, betray the operation of this secondary criterion by exhibiting an unexpected and impossible delicacy of discrimination.

reaction in the finger tip converts the test, virtually, into a test of discrimination of lifted weights.[1]

Treatment of Data.—The calculation of the difference limen and of the discriminative sensitivity is similar to that in the preceding test, save, of course, that the standard is now 100, instead of 80 g.

Results.— (1) *Normal capacity.* The discriminative sensitivity for cutaneous pressure depends so largely upon the type of the instrument (including especially the area of the pressure stimulus and the manner of application) that the norms obtained with other instruments can not be assumed to hold good for the present form of balance. Jastrow's results indicate a constant of approximately 1/15, which is nearly equal to that for lifted weights. Merkel similarly, reports 1/14 for his pressure balance, though Griffing believes that so fine a capacity as this must be attributed to the presence of a "muscular reaction of the finger." With a standard of 100 g. applied to the palm of the left hand, Miss Thompson found limens ranging from 4 to 20 g.

(2) *Dependence on the standard pressure.* The limen for the same *S*, with the same instrument and method, is constant, at least for stimuli between 50 and 2000 g. (Weber's Law).

(3) *Dependence on the area of stimulation.* According to Külpe (6, p. 160), the limen is 1/19 to 1/20 with an area of contact 1 mm. in diameter, but rises to from 1/13 to 1/16 with an area of contact 7 mm. in diameter.[2] Griffing, however, declares that "the area of stimulation does not, on the whole, affect the accuracy of discrimination for weights, but individual peculiarities appear in the results obtained."

(4) "*Practise* seems to aid discrimination at places not accustomed to pressure stimuli" (Griffing).

(5) There is no constant *sex* difference (Dehn, 2, and Thompson, 9).

(6) *Dependence on length of interval.* Accuracy of discrimination does not vary appreciably when the interval between application

[1] If it were not for the awkwardness of the position, it would, perhaps, be better to insert the finger volar side uppermost, in order more certainly to ensure against this movement of reaction in unreliable *S*'s.

[2] The tips of the author's balance are 8 mm. in diameter.

of the two stimuli is prolonged to 10 sec. (Griffing) or even to 30 sec. (Weber).

(7) *Dependence on place stimulated.* For weights of 100 g. or more, there is no appreciable difference in the discrimination of pressure on the palm of the hand, back of the hand, and the volar side of the index finger, though the last is probably more sensitive for very light weights (Griffing).

(8) *Constant error.* Most *S*'s show a tendency, frequently a marked tendency, to overestimate the second weight, *i. e.*, to judge it to be heavier (Griffing).

(9) *Direct judgments.* The impression of the standard stimulus not infrequently becomes so clear that it is carried over from one trial to another, so that, at least with large stimulus-differences, *S* may pass judgment when the first pressure is applied.

REFERENCES

(1) T. L. Bolton and Donna L. Withey, On the relation of muscle sense to pressure sense, in Nebraska Univ. Studies, 7: No. 2, April, 1907, 175-195.

(2) W. Dehn, Vergleichende Prüfung über den Haut- und Geschmack-Sinn bei Männern u. Frauen verschiedener Stände, Dorpat, 1894.

(3) J. A. Gilbert, Researches upon school children and college students, in Iowa S., 1: 1897, 1–39.

(4) H. Griffing, On sensations from pressure and impact, in P. R. M. S., i: 1895, No. 1, pp. 88. (Also Columbia Univ.Contr. to Phil. Psych. and Educ., iv.) Summarized in P. R., 2: 1895, 125-130.

(5) J. Jastrow, On the pressure sense, in A. J. P., 3: 1890, 54-6.

(6) O. Külpe, Outlines of psychology, Eng. tr., London, 1895.

(7) J. Merkel, Die Abhängigkeit zwischen Reiz. u. Empfindung (II), in Ph.S., 5: 1889, 245.

(8) E. C. Sanford, A course in experimental psychology, Boston, 1895 and 1898. Pp. 449.

(9) Helen B. Thompson, The mental traits of sex, Chicago, 1903. Pp. 188.

(10) E. B. Titchener, Experimental psychology, vol. ii., Quantitative experiments, N. Y., 1905.

(11) G. M. Whipple, New instruments for testing discrimination of brightness and of pressure and sensitivity to pain, in J. E. P., 1: 1910, 101-106.

TEST 22.

Sensitivity to pain.—The determination of the threshold or limen for pain has been conducted for the usual comparative purposes,

but it has had, in addition, a peculiar interest for some investigators, because it has been assumed that the limen varies in a characteristic manner with sociological status.

For the determination of the pain limen, use has been made both of electrical and of mechanical stimulation. Electrical stimulation (induction-coil current) has been employed chiefly by the Italian criminologists: pressure stimulation (upon the temple, palm, or finger-tip) has been employed almost exclusively by more recent investigators, and to this form of test our attention will be chiefly confined.

The value of the conclusions that have been so far reached from the use of pain tests is minimized by the difficulties, not always clearly realized, which appear in their administration. These difficulties, like those of most functional tests, arise primarily from the presence of a number of variable factors. The most important of these factors are: (1) dependence of the limen upon *S*'s ability to keep pain distinct both from strong pressure and from simple discomfort, (2) dependence upon the rate of application of the stimulus, including the length of time elapsing between successive applications, (3) dependence upon the place of stimulation, (4) dependence upon the area of the stimulus, (5) dependence upon the general condition and attitude of *S*, his "good-will," degree of fatigue, amount of practise, etc., (6) dependence upon individual constitutional differences in sensitivity, including sex, age, etc. This last is, of course, the particular dependence sought for in the results; the others, then, constitute disturbing factors and must, accordingly, be eliminated or at least evaluated. Proofs of these several dependences are given below in the discussion of results, but the first and second of them demand consideration here because they determine the choice of apparatus and of method.

(1) The dependence upon *S*'s judgment as to what constitutes pain has been recognized by most investigators as the primary source of difficulty in this test. It is undoubtedly true that pain is a specific sensory quality, distinct from pressure and distinct from unpleasantness; yet it appears equally true that it is often difficult, even for a practised *S*, to disentangle from his experience the three elements—cutaneous pressure, cutaneous pain, and discomfort. Children, to say the least, are not always competent to make such a differentiation, at least with the method of procedure that has commonly been followed.

To meet this difficulty, some *E*'s have instructed their *S*'s to wait for the distinct appearance of pain; others have asked them to report the first appearance of discomfort—an affective experience that might, or might not, be accompanied by pain, and which can thus scarcely be regarded as a rational index of the real pain limen. Thus MacDonald says (13, 14): "As soon as the subject feels the pressure to be in the *least disagreeable*, the amount of pressure is read from the scale . . . . The subject sometimes hesitates to say just when the pressure becomes the least disagreeable, but this is part of the experiment (!). The idea is to approximate as near as possible to the threshold of pain." Griffing, however, thinks that there is little liability to error from this source: "It is very easy to tell," he says "when the pressure begins to be uncomfortable, and the 'imagination' does not seem to be a disturbing factor. Indeed, the pain seems often to come with greater [great?] suddenness." We may wonder, then, why he says in the preceding paragraph: "The observers were asked to speak when the instrument began to hurt at all or to be uncomfortable; for it was found that individuals differed as to what they called 'pain.'" Perusal of the literature makes it evident, as these instances illustrate, that some *E*'s have been measuring a "discomfort" limen, other a "pain" limen: doubtless, in either case, some *S*'s reported "discomfort," while others reported real "pain."

Yet again, it appears that schoolboys have sometimes understood the test to be a measure of their endurance of pain, and have manfully asserted: "It doesn't hurt yet" when the pain limen has long been exceeded.

This difficulty of identification of the pain consciousness can not be wholly avoided, but it may be met, in part by giving *S* a clear account of the experience he is to report as pain, in part, especially in doubtful cases, by repeating the test, and in part by comparison of the results of a given *S* with the established norms for individuals of his age, sex, and type.

(2) Unless the rate of application of the pressure is constant from test to test, there is introduced a serious variable error. Roughly speaking, the limen will be higher if the pressure is applied rapidly, lower if it is applied slowly. In the use of the ordinary type of pain-tester, *e. g.*, of Cattell's (see 14, p. 1161) or of MacDonald's (12 and 14, pp. 1155-6) algometer, the rate of application of pressure is difficult to control; moreover, the rate has never been standardized, so that different investigators have followed different rates.[1] The chief merit of the pain-balance, or balance-algometer, as employed by Gilbert or as prescribed in the present test, consists in the guarantee that it affords of a rate of pressure increase that shall be uniform from step to step during each trial and from trial to trial.

Again, the time interval between successive trials must be standardized. If a given region, say the right temple, has been tested, its sensitivity is

[1] In illustration, Griffing applied pressure with the Cattell algometer at the rate of 1.4 kg. per sec., whereas Gilbert applied pressure with his own instrument at the rate of 50 g. per sec., or only $\frac{1}{28}$ as fast.

increased for some time thereafter, yet some investigators have not hesitated to make a series of 5 or 6 tests upon the same spot in immediate succession. On the other hand, if a given region be subjected to daily tests for several weeks, its sensitivity becomes reduced by a process of inurement. It is clear that both of these sources of error must be avoided in the determination of the limen.

The remaining sources of error are fully illustrated below, and are intelligible without further discussion. When all precautions have been taken, it is probable, however, that the results of this test will be more variable and less reliable than those of other psychophysical tests.[1]

APPARATUS.—The author's (20) pressure-pain balance (Fig. 47). Cardboard screen with suitable supports. Seconds' pendulum (Fig. 21) or other device for time-control. A low table. [Telegraph sounder (Fig. 26), battery and wire.]

PRELIMINARIES.—Arrange the instrument and screen as in Test 21. Place the pendulum within easy range of vision, or, since the time-relations are so important, convert it into an auditory signal by the use of the telegraph sounder and battery (preferably adjusted to give a rather faint click).

METHOD.—Seat *S* comfortably so that his hand lies upon the hand-rest with his finger-tip between the pressure-tips of the instrument, as described in Test 21. Give him the following instructions: "I want to measure your sensitiveness to pain. There is nothing for you to be afraid of, as I will stop the moment you tell me that you notice any pain. I shall add these weights, one after the other, on the end of this bar, and I want you simply to notice what you feel in your finger-tip. The pressure will grow stronger, bit by bit. It will, perhaps, feel uncomfortable after a time, but never mind that. Wait for the first moment when it really hurts, when you feel a stinging, sore feeling, or a real ache. Do you understand what I mean? I don't want to know *how much* pain you can 'stand' without crying out; I don't want to know when it is simply

[1] The comments just given make it evident that MacDonald's algometer and Cattell's algometer are inadequate instruments. It follows that the results published by Griffing, Wissler, Swift, MacDonald, and Miss Carman are of doubtful value. Gilbert's results, though obtained by a better instrument, are more uneven than those of any other of the tests that he undertook and, on account of the slow rate of application that he used, are not directly comparable with those obtained by the method outlined below.

uncomfortable; I want to know when you first notice what you would call *actual pain*."

Throw the release-lever up to the left, so that the support beneath the balance-beam is permanently depressed: this makes possible a continuous, but cumulative pressure upon the finger.

Apply one of the large brass disc-weights, marked *B-200 g.*, every 2 sec. These discs are placed, *without jar*, upon the pin marked *B*, at the outer end of the beam. Continue application until *S* reports pain, then immediately remove the pressure. The total weight on the beam measures *S*'s pain limen.

Repeat the test with the left forefinger.

VARIATIONS OF METHOD.—(1) Test other fingers of both hands.

(2) Apply the stimulus weights at a slower rate, say once in 4 sec., and note the effect upon the limen.

(3) Substitute a series of pressures for the cumulative, continuous pressure, by using the release-lever as in Test 21, and applying the pressure for 1 sec. only, after each weight is added.

RESULTS.—(1) The short series of tests available with the pressure-pain balance thus far has shown that, with adults, the pain limen may be expected to lie between 1600 and 2400 g., when the method of cumulative pressure is employed.

(2) *Dependence on rate of application.* (*a*) "The rate at which pressure is added influences greatly the amount that is required to produce pain" (Gilbert).

(*b*) Immediate or close repetition of stimulation causes increased sensitivity, but continued practise for several weeks appears to reduce it (Griffing, 8).

(3) *Dependence on nature of stimulus.* Sensitivity to pain produced by electrical stimulation bears no noticeable relation to that produced in the same person by pressure stimulation (Griffing, 7).

(4) *Dependence on the area of the stimulus.* "The pain threshold increases with the area of stimulation in an approximately logarithmic proportion" (Griffing, 8). Thus, when Cattell's algometer was applied to the palm of the hand, areas of 10, 30, 90 and 270 sq. mm. were correlated with limens of 1.4, 2.8, 4.4, and 6.6 kg., respectively.[1]

[1] The area of the pressure tip in the author's balance is approximately 50 sq. mm.

(5) *Dependence on region.* The regions of the body most sensitive to pain (from pressure stimulation) are those over the frontal and temporal bones, while the heel, the back, and the muscular regions of the leg and the hand are distinctly less sensitive. Illustrative limens obtained by Griffing are:

	KG.		KG.
Top of head	= 1.8	Right thigh, ventral surface	= 4.3
Forehead	= 1.3	Left hand, volar side	= 6.2
Right temple	± 1.0	Right heel, plantar side	= 7.0
Left temple	= 1.3	Back	= 8.0

These differences seem "to depend largely upon the thickness of the skin and the extent of the subcutaneous tissues." The left side of the body is in general somewhat more sensitive than the right.

(6) *Dependence on sex.* Gilbert (6), MacDonald, Dehn (4), Carman (2), Swift (16), and Wissler (21) agree that women are more sensitive to pain than are men. Thompson (17) agrees to this generalization, but adds that there are more men than women with very low thresholds, *i.e.*, that there is greater variability in men. Ottolenghi (15) and Lombroso (10), on the other hand, state that (with electrical stimulation) women are markedly less sensitive than men.

The latter authority believes that this result is confirmed by the experience of surgeons, who find that women possess greater endurance of pain the popular opinion that women are more sensitive to pain is due, in his view, to the greater tendency of women to express feelings of pain by tears or otherwise; he also believes that their greater longevity may be due partly to their inferior susceptibility to pain.

Typical results are those of Wissler and of Gilbert: the former publishes the following averages (Cattell algometer on the ball of the right thumb); college men, 5.9 kg.; college women, 2.4 kg. Gilbert finds that the average difference between boys and girls is about 400 g., and that this difference increases with age, until at 18 or 19 it becomes over 1 kg. (See Table 28.)

(7) *Dependence on age.* Sensitivity to pain, in general, decreases with age up to 18 or 19 years, and is thenceforth approximately stationary, but Carman and MacDonald both find irregularities near the period of puberty, and Wissler finds Seniors more

sensitive than Freshmen (as a class). It is probable that the general result is disturbed by a tendency on the part of younger children to shrink from the test and to report discomfort rather than pain. Gilbert's results are embodied in Table 28.[1]

TABLE 28

Pain Limen, in kg., for about 50 Boys and 50 Girls of each Age (Gilbert)

AGE	6	7	8	9	10	11	12	13	14	15	16	17	18
Boys	1.26	1.38	1.70	1.69	1.67	2.07	2.00	2.05	2.13	2.35	2.70	2.75	2.85
Girls	1.15	0.93	1.18	1.36	1.45	1.56	1.46	1.70	1.82	1.77	1.85	1.93	1.80
Average	1.21	1.16	1.44	1.53	1.56	1.82	1.73	1.88	1.98	2.06	2.28	2.34	2.33

(8) The range of *individual difference* is large in the pain test. Gilbert, for example, found that the mean variation for his groups of children ranged from 330 to 820 g.

(9) *Dependence on fatigue.* Swift (16) concludes that fatigue increases sensitivity to pain, especially in the case of younger pupils and of girls, because it lowers the tone and increases the irritability of the whole system.[2] Essentially similar results are reported by Vannod (18) and by Vaschide (19). The former employed an instrument ('algesiometer') analogous to v. Frey's 'hair-esthesiometer,' and found that the pressure needed to produce pain fell from 45 g. at 8 a.m. to 39 g. at 10 a.m., and to 29 g. at 12 m., under the influence of school work. The latter concluded that pain tests warranted him in stating (1) that mathematics and ancient languages possess an especially high fatigue-value, (2) that written exercises in the form of tests produce intense intellectual fatigue, (3) that afternoon is much more fatiguing than forenoon instruction, and (4) that a forenoon spent outside of the school permits

[1] It should be remembered, again, that Gilbert used a very slow rate of application, so that his results may not be comparable with those obtained by the methods we have prescribed.

[2] This conclusion is based upon tests before and after a 10 days' vacation in which the "physical condition" was determined by a dynamometer—a method already shown to be of doubtful value (Tests 6 and 9). There is no evidence to indicate that check-tests were made to determine the range of variations that might have appeared under constant work-conditions.

a return, in most cases, to normal sensitivity. On the other hand, Binet (1), who used an adaptation of Blocq's sphygmometer, has come to diametrically opposite conclusions, and asserts that the effect of fatigue is to reduce, not to heighten, pain sensitivity.

(10) *Dependence on mental ability.* The relation of the pain limen to mental ability is not clear. Carman found that bright boys (teacher's estimate) were more sensitive than dull boys: Swift confirmed this by contrasting the best with the poorest fifth of a class, and attributes the result to the more delicate nervous organization of bright children. MacDonald, however, says "there is no necessary relation between intellectual development and pain sensitiveness." "Obtuseness to pain seems to be due more to hardihood in early life."

A curious and somewhat dubious correlation unearthed by Miss Carman is that boys and girls who are especially dull in mathematics are more sensitive on the right than on the left temple.

(11) *Dependence on sociological condition.* (*a*) Similarly unconvincing is the series of conclusions in which MacDonald (11, 13, 14) summarizes his correlations between pain sensitivity and sociological condition; *e.g.*, girls in private schools, who are generally of wealthy parents, are more sensitive than girls in public schools; university women are more sensitive than washerwomen, but less sensitive than business women; self-educated women are more sensitive than business or university women (owing, perhaps, to overtaxing their nervous systems in the unequal struggle for an education); the non-laboring classes are more sensitive than the laboring classes, etc.[1]

(*b*) The study of the pain sensitivity of the *criminal* is a specific sociological problem that has attracted much attention since the concept of the 'criminal type,' or of the 'instinctive criminal,' was introduced by Lombroso and his school. It has been generally stated that the typical criminal is distinctly less sensitive to pain than the average normal man, and it has frequently been added that the normal insensibility of the criminal is to be largely attributed to

[1] The measurements from which these conclusions are drawn were made by different investigators, by an unreliable method, and have been assembled apparently by mere comparison of averages and with no attempt to determine the limit of error; they might, or might not, be confirmed by more exact methods.

this bodily insensibility. These statements are based upon certain experimental tests and upon common observations of the hardihood and general obtusity of the 'typical criminal.' Nevertheless, recent pain measurements indicate that the generalization is too sweeping, and that numerous exceptions occur. It may even be doubted whether the existence of a distinct criminal type has been satisfactorily established.

A general summary of the pain sensitivity of criminals is given by Ellis (5, Section 8). The inadequacy of the algometer test as applied to criminals is discussed briefly by Miss Kellor. A typical exception to the general belief is found in Dawson's conclusion (3) that normal children are less sensitive to pain than delinquent children, probably because many of the delinquents were of neurotic type.

(12) *Miscellaneous correlations* reported by Miss Carman are: boys with light hair and eyes are less sensitive than boys with dark hair and eyes. First-born are more sensitive than second-born boys, and the latter than later-born brothers: the same is true of girls, save on the right temple (!). These conclusions are subject to obvious criticism.

REFERENCES

(1) A. Binet, Recherches sur la fatigue intellectuelle scolaire et la mesure qui peut en être faite au moyen de l'esthésiomètre, in A. P., 11: 1905, 1-37, especially 32 ff.

(2) Ada Carman, Pain and strength measurements of 1507 school children in Saginaw, Michigan, in A. J. P., 10: 1899, 392-8.

(3) G. E. Dawson, A study in youthful degeneracy, in Pd. S., 4: 1896, 221-258.

(4) W. Dehn, Vergleichende Prüfung über den Haut- und Geschmack-Sinn bei Männern u. Frauen verschiedener Stände, Dorpat, 1894.

(5) H. Ellis, The criminal, 3d ed., London, 1907.

(6) J. A. Gilbert, Researches upon school children and college students, in Iowa S., 1: 1897, 1-39.

(7) H. Griffing, On individual sensitivity to pain, in P. R., 3: 1896, 412-5.

(8) H. Griffing, On sensations from pressure and impact, in P. R. M. S., 1: 1895, No. 1. Pp. 88 (also Columbia Univ. Contr. to Phil., Psych. and Educ., 4). Summarized in P. R., 2: 1895, 125-130.

(9) Frances Kellor, Experimental sociology, N. Y., 1901. Pp. 316.

(10) C. Lombroso, The sensibility of women, brief report in Mind, n.s. 1: 1892, 582.

(11) A. MacDonald, Sensibility to pain by pressure in the hands of individuals of different classes, sexes and nationalities, in A. J. P., 6: 1895, 621-2.

(12) A. MacDonald, A temporal algometer, in P. R. 5: 1898, 408-9.

(13) A. MacDonald, Further measurements of pain, in P. R., 6: 1899, 168-9.

(14) A. MacDonald, Experimental study of children, etc., reprint chs. 21 and 25 of U. S., 1897-8, Washington, 1899.

(15) S. Ottolenghi, La sensibilité de la femme, in Revue scient., Ser. 4, vols. 5: 395, and 6: 698.

(16) E. Swift, Sensibility to pain, in A. J. P., 11: 1900, 312-7.

(17) Helen B. Thompson, The mental traits of sex, Chicago, 1903. Pp. 188.

(18) Th. Vannod, La fatigue intellectuelle et son influence sur la sensibilité cutanée, in Rev. méd. de la Suisse Romande, 27: 1896, 21.

(19) N. Vaschide, Les recherches expérimentelles sur la fatigue intellectuelle, in Revue de philos., 5: 1905, 428.

(20) G. M. Whipple, New instruments for testing discrimination of brightness and of pressure and sensitivity to pain, in J. E. P., 1: 1910, 101-106.

(21) C. Wissler, The correlation of mental and physical tests, in P. R. M. S., 3: No. 6, 1901. Pp. 62.

TEST 23

Discrimination of dual cutaneous impressions: Esthesiometric index.—As long ago as 1834, E. H. Weber, a German physiologist, observed (50) that, if two punctiform pressures are applied simultaneously to adjacent points on the skin, a single impression results, whereas, if the pressure points are applied at gradually increased distances, an extent can be discovered which is just sufficient to yield a perception of two points. Weber explored many regions of the skin and published extended tables of measurements of this distance, which has since become known variously as the "limen for duality," or "doubleness," as the "esthesiometric index," the "space threshold," or even, less exactly, as the "index of delicacy of touch."

On account of Weber's explanation of the phenomenon, which was in terms of the supposedly quasi-circular distribution of the end-organs of the sensory nerves, the experiment is often referred to as the test of "sensory circles." On account of the type of instrument employed, it is sometimes termed the "compass test."

Since Weber's time the experiment has become a classic in psychology. Seemingly simple and definite, more careful examination

has revealed the fact that the determination of the esthesiometric index is in reality unusually difficult, and that the factors which underlie the observer's judgment are surprisingly varied and subtle.

For differential psychology, the chief interest in the test is found in its use by criminologists to measure "general sensibility," and by several German investigators to measure the degree of fatigue of school children. Physicians, also, have employed it for diagnostic purposes, particularly in connection with pathological conditions of the spinal cord, and it has found special favor in the psychological laboratory, both for its intrinsic interest and for the illustration of various psychophysical methods.

As in the case of other tests, the chief difficulty in the use of the esthesiometric test lies in the presence of numerous sources of error, which must be fully recognized and controlled if valid results are to be secured. In general, it may be said that the esthesiometric limen will depend upon (1) the instrument employed, (2) the region of the body tested, (3) the method of procedure, including the nature of the instructions, (4) the care with which *E* applies the stimulus and the actual pressure employed, (5) *S*'s degree of fatigue, (6) *S*'s degree of practise, (7) *S*'s ability to attend to the impressions and to make accurate reports, especially in the 'critical region,' and (8) upon a number of other factors, such as *S*'s sex, age, the condition of the circulation in the region tested, etc. The manner in which these factors affect the index will be discussed below.

The instrument employed may be extremely simple, *e.g.*, the set of needles thrust through bits of cardboard, used by Binet in his earlier tests (2), or it may be very complicated and elaborate.

In general, the development of the esthesiometer since Weber's time has been in the direction of greater complexity and delicacy, with a view of affording more adequate control of the separation of the points, of the simultaneity of their application, and of the degree of pressure exerted. It is doubtful whether much of this elaboration is needful: objective equalization of the pressure does not insure subjective equalization, and a careful *E* is better able to apply the points simultaneously if he works with a relatively simple instrument.

The instrument selected, an improved form of Jastrow's esthesiometer, possesses all the requisite features. For other models, consult Blazek (7), Binet (3, 4), and Washburn (48). The models of Ebbinghaus and v. Frey are figured in Zimmermann's catalog. Spearman's instrument is described in Sommer (38) and pictured in use in Schulze (33, p. 67).

Apparatus.—Jastrow's improved esthesiometer (Fig. 48). Cardboard screen and supports. Pillow or folded towel.

Method.—(*a*) *Preliminary practise.* Seat *S* comfortably with his right forearm laid horizontally, volar side uppermost, upon a small pillow or folded towel, with the clothing arranged to expose the forearm from elbow to wrist, without impeding the circulation at the elbow.

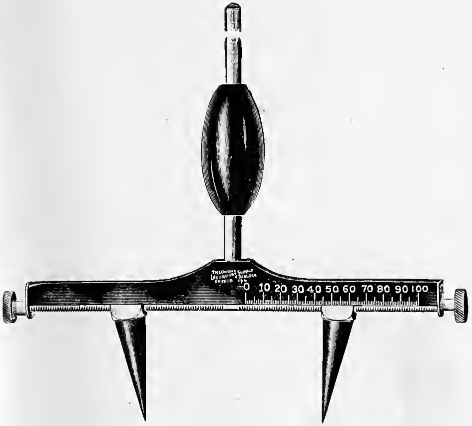


FIG. 48. JASTROW'S IMPROVED ESTHESIOMETER.

Arrange the screen to cut off from *S* the view of his forearm and of the instrument.

Devote from 2 to 5 min. to a preliminary practise series in order to familiarize *S* with the test, particularly with the perception of one and of two points.

Instruct all *S*'s, in the same words, substantially as follows: "I'm going to touch your arm with points, something like pencil-points. They won't hurt you at all. You are to give careful attention to what you feel, and tell me if you think I'm touching you with one point or with two points. You will have to watch very carefully. If you feel only one point, say "one:" if you feel two, say "two."

Begin with a distinctly supraliminal distance, say 90 mm. After

a warning 'ready' signal, bring the instrument down perpendicularly upon the middle of the forearm, parallel with its longitudinal axis, in such a manner that both points make contact simultaneously,[1] and rest by their own weight upon the skin, as the holder is allowed to slide one or two cm. down the stem of the instrument. The several applications should be uniform in duration—about 1.5 sec.

Next apply a single point, then two points at 80 mm., and so on, giving sometimes one point and sometimes two points, of varying distances and in irregular order. Leave an interval of 7 to 10 sec. between trials to allow the preceding sensory disturbance to die away. Work as nearly as possible in the same region, but to avoid local fatigue, do not seek to apply the compass upon exactly the same spots at each trial. Avoid contact with hairs, which set up tickling sensations, or pressure upon projecting veins or tendons, which will be of a character dissimilar to the normal contact.

This practise should help *S* to be familiar with, and to distinguish the 'feels' of one point and two points. It will be found that some *S*'s will, nevertheless, occasionally answer "two" when but one point is given. This is the not uncommon *Vexirfehler*, or esthesiometric paradox, which is well recognized as a source of difficulty in esthesiometry. If it proves persistent, it will probably be impossible to determine an exact limen with that particular *S*. To avoid it, *E* may allow *S* to look at the instrument when one point is resting on the skin and he has just announced "two." Sometimes young *S*'s, who have caught sight of the instrument, may, with childish logic, conclude that there must always be two points because there are two on the instrument.

Children who display timidity must be encouraged to adopt a more favorable attitude toward the test.

(*b*) *Test proper*. Allow *S* a short rest (during which his arm should be withdrawn from the somewhat constrained position); then resume the experiment in accordance with the general instructions for discrimination work given at the beginning of the chapter. Proceed from a distinctly supraliminal distance toward the critical region until a distance is found that proves difficult for *S* to judge. Then give a series of 10 double contacts at that distance, but add to these, two or three single contacts for check tests. Make a

[1] This is imperative because the limen for successive stimuli is only $\frac{1}{3}$ to $\frac{1}{4}$ that for simultaneous stimuli.

record of the order used, and follow it for each S. If S makes 10 correct judgments for the 10 double contacts, try another series with a lesser separation: if he makes but 5 or 6 correct judgments, increase the separation. Seek a distance such that about 8 correct judgments in 10 are made, *i.e.*, such that double contact is reported as double in 8 of 10 trials. The check tests should, of course, be judged "one," but if the paradoxical error does appear, E must either try a wider separation, or give up, for the time being, the attempt to find a *definite* limen.

Variations of Method.—Test other regions of the body. Use a transverse application on the forearm. Compare the sensitivity before and after periods of rest or of fatiguing work. Test the acquisition of practise and its transfer to symmetrical and adjacent portions of the body.

Results and Conclusions.[1]—(1) *Dependence on sex.* Wissler (51) could discover no sex differences in cutaneous discrimination of dual pressure, but Miss Thompson (41), despite some acknowledged difficulties in the administration of her test, concludes that "women have a somewhat finer discrimination in the crosswise direction, and a decidedly finer discrimination in the lengthwise direction." Actual figures cited are 20 and 65 mm. for women and 35 and 75 mm. for men, in the transverse and longitudinal directions, respectively.

(2) *Dependence on age.* There is fair agreement, especially among the earlier investigators (see *e.g.*, Czermak, 11) that children have a greater sensitivity than adults. Similarly, Wissler found Seniors inferior to themselves as Freshmen. In regions where the sensitivity is poor, the difference between children and adults is quite distinct, *e.g.*, a limen of 67 mm. on the thigh of an adult in contrast to 35 mm. upon the same region in a boy of 12.

The usual explanation that the child has a greater number of nerve endings within the same sized area is probably only a partial explanation, as it has been computed that the child's sensitivity is out of all proportion to the differences in dimensions here concerned. A contributory factor is doubtless to be found in the fact that the child's skin is thinner and more tender, so that a given impact produces a sharper sensory experience.

---

[1] In reporting these results, it may be stated that many of them have been obtained by methods that are open to criticism, particularly in that suitable precautions have been wanting to control or to eliminate the numerous disturbing factors already mentioned.

Griesbach (15), who found no difference between children from 11 years up to 19, is practically the only investigator who has not found the child more sensitive than the adult, but Griesbach's figures can not be accepted without misgivings.

(3) *Dependence on region.* (*a*) Table 29, which is derived from Weber's original results, gives an idea of the topographic distribution of sensitivity.

TABLE 29

Topography of Esthesiometric Sensitivity (Weber)

REGION	LIMEN IN MM.	REGION	LIMEN IN MM.
Tip of the tongue	1.1	Forehead	22.5
Tip of the fingers	2.2	Back of hand	31.5
Mucous membrane of the lips	4.5	Forearm	40.5
End of the nose	6.7	Back	54.1
Cheek	11.2	Thigh	67.6

An empirical generalization, known as Vierordt's Law (45, p. 298), summarizes these differences in sensitivity, especially of the limbs and head, by the statement that the delicacy of discrimination of two regions on the skin of a portion of the body that is moved as a whole is proportional to the average distance of these regions from their common axis of rotation.[1] In illustration, if the discriminative sensitivity of the tip of the shoulder (acromion) be taken at 100, then that of the upper arm is 151, of the forearm 272, hand 659, thumb 2417, middle finger 2582.

(*b*) Van Biervliet (42), who asserts that in many sense-departments the favored *side of the body* is superior to the other side by the fraction $\frac{1}{9}$, publishes tables which indicate that the same constant applies in esthesiometry.

(4) *Dependence on time of day.* A comparison of the sensitivity at different periods of the day, in the search for a diurnal rhythm, is, of course, complicated by the presence of fatigue (see the

[1] Somewhat analogously, Krohn (25) cites an instance in which a man whose arm had been held immovable in a plaster cast for three months exhibited marked decrease of sensitivity of that arm shortly after the cast had been removed.

following paragraph), and doubtless by individual differences as well. Schmey (32) believes sensitivity to be less at night than in the morning. Adsersen (1), however, asserts that sensitivity is lowest in the morning, begins to improve at about 11, reaches a maximum from 3 to 7, and thence decreases. By comparison, he shows that this rhythm coincides very closely with the diurnal curve of bodily temperature, and he therefore argues that the esthesiometric fluctuations are indices of physiological changes common to the diurnal rhythm. Tawney's work (39), on the other hand, leads one to believe that the limen undergoes, in many *S*'s, such irregular fluctuations that it is impossible to find a limen that is constant for half an hour at a time.

(5) *Dependence on fatigue.* The effect of fatigue upon the limen forms the chief source of interest in connection with the esthesiometric test. Griesbach (15) was the first to make extended use of the test in the examination of school children. His amazingly uniform and definite results, when taken at their face value, indicate unequivocally that the method is of value and importance in the detection of the fatigue induced by school work. They have, moreover, been confirmed more or less thoroughly by Binet (6b), Blazek (7), Bonoff (10), Heller, Schuyten (35), Vannod (44), and Wagner (47). On the other hand, they have been controverted with equal emphasis by Bolton (8, 9), Germann (13), Kraepelin (24), Leuba (25), and Ritter (30). Investigations of fatigue by means of the esthesiometric test have also been made of late by Ferrai in Italy, Sakaki (31) in Japan, Ley in Belgium, Michotte in Belgium, and Noikow (28) in Bulgaria.[1]

Griesbach argues that fatigue reduces the power of sustained attention' and that this in turn reduces the cutaneous sensitivity. He appears to believe that he is the first to have discovered this relation, but it may be noted that Weber, himself, had cautioned his readers to avoid fatigue if valid results were to be secured, that Schmey, in 1884, had demonstrated that the fatigue of the arm by calisthenic exercises reduced sensitivity, and that Stanley Hall (17), in 1879, had commented on the variability of the results obtained from Laura Bridgman, and had expressed the hope that "a curve

---

[1] For further discussion of the value of the method, consult Gineff (14), Meumann (27, vol. ii., 89–94, 107–110) and the recent monograph of Offner (29, 18–23). Griesbach's rejoinder to his critics (16) should also be consulted.

of fatigue may be obtained by which some approximate comparison with the fatigue of a nerve-muscle preparation may be made." Griesbach worked on the glabella, cheek-bone, tip of the nose, under lip, ball of the thumb, and tip of the index-finger, and tested pupils before and after various kinds of school work, on Sundays, holidays, at the end of vacations, etc. A single example will suffice: a girl of 14 had a limen on the glabella of 5 mm. at 7 a.m., but this increased to 12.5 mm. at 12, noon, after a morning at school, whereas on Sunday her limen was but 3.5 mm. Griesbach concludes his memoir with a strong plea against over-work, and asserts that no schoolboy can meet in full the demands of present-day higher education without endangering his health (15, p. 88).

Blazek divides pupils into three types, (1) those who possess ability, who work industriously and attentively, and thus exhibit distinct and progressive fatigue curves, (2) those who work intermittently, and whose curve is therefore broken by recuperative periods, (3) those whose curve is approximately a straight line. The fatigue curve, therefore, depends in the main upon the type of worker, but the individuality of the teacher and the subject-matter in hand are also determining influences. He concludes that more than half the pupils work irregularly and thus save themselves in part from over-work. He would recommend 4 subjects daily, of 45 min. each, with 15 min. rest-periods between each subject.

Wagner says that the Griesbach method is a valuable adjunct for the study of fatigue. Afternoon instruction is practically valueless pedagogically. Play and gymnastics are sources of fatigue to many pupils, and should be relegated to the close of instruction or to the afternoon. If the fatigue-value of mathematics be placed at 100, other subjects may be rated thus: Latin 91, gymnastics 90, geography and history 85, French and German 82, nature-study 80, drawing and religion 77.

Binet summarizes the results obtained by himself, and by a group of teachers who worked under his direction, by declaring that "intellectual fatigue is manifested by a reduction of sensitivity, measurable on the back of the hand: this reduction is revealed by fewer judgments of 'two' for smaller distances (0.5 to 1.5 cm.), is more pronounced for girls than for boys, and is to be attributed to an actual reduction of tactual sensitivity itself, not to a mere relaxation of attention" (6 b, p. 29). It is to be noted, however, that this conclusion is based upon the 'lump' results obtained from groups of school children: analysis shows that, in some groups at least, less than half of the pupils (*e.g.*, 31 of 75), gave evidence of fatigue in this way. Moreover, Binet's method of distinguishing between the effects of lessened sensitivity and of lessened attention is open to debate.

Schuyten at first condemned the esthesiometric method (34), but later (35) he found to his surprise that it worked satisfactorily. He used groups of 5 selected pupils.

Ritter tested himself at intervals for two years, but could get no evidence of fatigue by the esthesiometer.

Bolton says the limen is so hard to determine that it can not be satisfactorily accomplished in a single sitting. In his tests, severe mental work of two hours duration did not produce a measurable change in the index.

Germann, after making 2450 trials on a single *S*, could not discover any relation between fatigue and the limen.

Leuba admits that fatigue affects the limen, but says that it is only one of a great many factors. He was unable to draw any general inductions concerning fatigue, even when the results for three days of severe mental work were contrasted with those for three days of rest.

Meumann (27, vol. 2, 90 ff.), similarly, admits that the tendency of fatigue is to heighten the limen, but protests vigorously that the relation is but indirect, and so complicated by numerous little-known factors that the numerical expression of shifts in the limen can in no wise be regarded as a measure of fatigue. ("Wir haben in der Erhöhung der Raumschwelle durch die Ermüdung *nur ein objectives Symptom derselben, aber keine Messung.*")

Kraepelin, however, declares flatly that investigations that embody measurements of fatigue by the use of the esthesiometer are "all in the air" (stehen einfach in der Luft), and are nothing but the unintentional expression of the preconceived opinions of the investigators.

(6) *Dependence on practise.* The effect of practise was studied by Czermak (11), as early as 1855, in his investigation of the sensitivity of the blind, was measured more carefully by Volkmann (46), and confirmed later in particular by Dresslar (12). From these investigations, it would appear that the practise-effect is visible within two hours, and may be pushed to unexpected lengths by continued work; thus, Dresslar reports one *S*, who started with a limen of 29 mm., reduced this to 21 mm. in the first week, to 10 mm. the second week, 5.5 mm. the third week, and 2.8 mm. the fourth week—a net reduction, then, to approximately $\frac{1}{10}$ of the original figure. This practise-effect is, however, rapidly lost, being reduced very definitely within 8 days and completely lost within a month. The practise-effect is said to appear much more rapidly on fingers, hands, and other exposed parts, than on the back and other relatively inaccessible and immobile regions. Both Volkman and Dresslar submit evidence to show that this 'education' is subject to *transfer to symmetrical regions,* though not to regions adjacent to the one on which it was effected.

In seeking an explanation for the effect of practise, we are met with the fact that other investigators have not confirmed the results just cited. Camerer, in his lengthy series, did not find such extensive practise-effects

Tawney (40) found his work so vitiated by *Vexirfehler* and by auto-suggestion in general that he was unable, for some S's, to establish any constant limen. Both Tawney and Henri (18, b) deny that the influence of practise, when established, is confined to symmetrical regions of the body. Solomons (37) asserts that the practise-effect is rapid if S is informed of his errors, but practically non-existent if he is not—a principle, which, if confirmed, may explain the disagreements just cited.

The explanation of the process of 'education' in this test is found by some writers to lie in certain peripheral processes—not necessarily in anatomical, since the practise is too rapid to admit of that, but rather in local physiological, processes. Others, *e.g.*, Judd (21) and Solomons (37), believe that the education is essentially a 'central,' or psychological, process—an improvement in judgment due to the learning of new associations.

The low *threshold of the blind*, reported by Goltz, Gärtner, Heller, Miss Washburn (49), Hall (17), Jastrow (10), and others, is to be deemed a special example of practise, and does not imply the presence of exceptional sensitivity or special peripheral delicacy. Helen Keller (20) has a limen of 1.5 mm. on the tip of the left forefinger, and 3-4 mm. on the palm of the hand, which is smaller than that of the average S. Laura Bridgman (17) is credited with a limen of 0.7 mm. on the right forefinger, and her general sensitivity of touch, according to Hall, was "from two to three times as great as that of an ordinary person."

(7) *Dependence on expectation.* An important feature of esthesiometry has been emphasized by the work of Solomons and of Tawney, both in the latter's own investigation (40) and in that undertaken with Henri (19). These writers have shown that expectation or *suggestion* plays a preponderant rôle in esthesiometry. By arranging the conditions suitably, they were able, for example, to induce practised and reliable S's to judge, with considerable uniformity, "one" when two points were given, and "two" when one point was given.[1]

It is of interest herewith to note that the esthesiometric paradox often appears more frequently as practise continues, and is more likely to appear in tests of adults than in tests of children.

(8) *Dependence on circulation of the blood.* The condition of the circulation in the region under test affects the limen. It appears

[1] These observations, which might be repeated in many other fields, show how essential it is to work methodically and under constant conditions. The extent to which S's discrimination is affected by his attitude toward the experiment, and by his manner of judging in general, has led Binet to declare (5) that the "compass test" measures "tactual intelligence" rather than the fineness of touch itself.

from the studies of Brown-Sequard, Schmey, and others, that arterial hyperemia increases cutaneous sensitivity, whereas aremia, or venous hyperemia, or decided cold, reduces sensitivity. Excessive stretching of the skin decreases sensitivity, as does the use of narcotics.

(9) *Dependence on intelligence.* (*a*) There is no clear relation between general intelligence and the esthesiometric limen. The rather crude test of Wissler (45) revealed no correlation with the class standing of university students. Binet (2) found his intelligent superior to his unintelligent group at the first trial, but the difference soon lessened as the boys became adapted to the test. Van Biervliet (43) used the compass test to secure a measure of intelligence, not by the limen itself, but by its mean variation, on the assumption that this latter measure is, in almost any test, the real index of intelligent work. His figures give the 10 most intelligent of 300 university students an index of 17.7, and the 10 least intelligent an index of 27.6. His index, however, as Binet (6 a) has pointed out, is a rather dubious device.

(*b*) As might be expected, the use of the test for the examination of *abnormal children,* criminals, truants, etc., is beset with difficulty, because it exacts prolonged, sustained attention and interest (Kellor, 22; Kelly, 23). Simon (36) was unable to test the lowest grade of children in a school for the feeble-minded, but obtained results from the less defective types which indicated that their sensitivity was less than that of normal children. Whether the difference is reducible to differences in cutaneous sensitivity itself, or to differences in ability to control the attention and understand directions, is not clear.

REFERENCES

(1) H. Adsersen, Eine aesthesiometrische Untersuchumg, in Z. S.. 27: 1904, 540-3.

(2) A. Binet, Attention et adaptation, in A. P., 6: 1899 (1900), 248-404.

(3) A. Binet, (*a*) Un nouvel esthésiomètre, in A. P., 7: 1800 (1901), 231-9; (*b*) Technique de l'esthésiomètre, 240-8.

(4) A. Binet, La mesure de la sensibilité, in A. P., 9: 1902 (1903), 79-128. Also five other articles on esthesiometry, pp. 129-252.

(5) A. Binet, De la sensation à l'intelligence, in Rev. philos., 56: 1903, 450-467, and 592-618.

(6) A. Binet, (*a*) A propos de la mesure de l'intelligence, in A. P., 11: 1905, 69-82. (*b*) Recherches sur la fatigue intellectuelle scolaire et la mesure qui peut en être faite au moyen de l'esthésiomètre, do., 1-37.

(7) B. Blazek, Ermüdungsmessen mit dem Federaesthesiometer an Schülern des Franz-Joseph-Gymnasiums in Lemberg, in Z. P. P., 1: 1899, 311-325.

(8) T. Bolton, The reliability of certain methods for measuring the degree of fatigue in school children, in P. R., 7: 1900, 136-7.

(9) T. Bolton, Ueber die Beziehungen zwischen Ermüdung, Raumsinn der Haut und Muskelleistung, in P. A., 4: 1902, 175-234.

(10) N. Bonoff, Étude médico-pédagogique sur l'esthésiométrie et la simulation a l'école, in I. M. S. H., 4: 1907-8, 384-394.

(11) J. N. Czermak, (*a*) Beiträge zur Physiologie des Tastsinnes, Wien., 15: 1855, 466-520; (*b*) Weitere Beiträge, etc., *ibid.*, 28: 577-600.

(12) F. B. Dresslar, Studies in the psychology of touch, in A. J. P., 6: 1894, 313-368.

(13) G. Germann, On the invalidity of the esthesiometric method as a measure of mental fatigue, in P. R., 6: 1899, 599-605.

(14) D. Gineff, Prüfung der Methoden zur Messung geistiger Ermüdung, Zürich, 1899. Pp. 68.

(15) H. Griesbach, Energetik u. Hygiene des Nerven-Systems in der Schule, Munich u. Leipzig, 1895. Pp. 97. Also published as Ueber Beziehungen zwischen geistiger Ermüdung u. Empfindungsvermögen der Haut, in Arch. f. Hygiene, 24: 124.

(16) H. Griesbach, Weitere Untersuchungen über Beziehungen zwischen geistiger Ermüdung u. Hautsensibilität, in I. M. S. H., 1: 1905, 317-417.

(17) G. S. Hall, Laura Bridgman, in Mind, 4: 1879, 149-172, especially 160 ff.

(18) V. Henri, (*a*) Revue générale sur le sens du lieu de la peau, in A. P., 2: 1895 (1896), 295-362 (with bibliography). (*b*) Ueber die Raumwahrnehmungen des Tastsinnes, Berlin, 1898. Pp. 228. (With extended bibliography).

(19) V. Henri and G. Tawney, Ueber die Trugwahrnehmung zweier Punkte bei der Berührung eines Punktes der Haut, in Ph. S., 11: 1895, 394-405.

(20) J. Jastrow, Psychological notes on Helen Keller, in P. R., 1: 1894, 356-362.

(21) C. Judd, Ueber Raumwahrnehmungen im Gebiete des Tastsinnes, in Ph. S., 12: 1896, 409-464. (Historical survey, 451-8).

(22) Frances Kellor, Experimental sociology, N. Y., 1901. Pp. 316.

(23) R. L. Kelly, Psychophysical tests of normal and abnormal children; a comparative study, in P. R., 10: 1903, 345-372.

(24) E. Kraepelin, Ueber Ermüdungsmessungen, in A. G. P., 1: 1903, 9-30.

(25) W. Krohn, Sensation-areas and movement, in P. R., 1: 1894. 280-1.

(26) J. Leuba, On the validity of the Griesbach method of determining fatigue, in P. R., 6: 1899, 573-598.

(27) E. Meumann, Vorlesungen zur Einführung in die exp. Pädagogik, Leipzig, 1907.

(28) P. Noikow, Aesthesiometrische Ermüdungsmessungen, in I. M.S.H., 4: 1907–8, 437–481.

(29) M. Offner, Die geistige Ermüdung, Berlin, 1910. Pp. 88.

(30) C. Ritter, Ermüdungsmessungen, in Z. P., 24: 1900, 401–444.

(31) Y. Sakaki, Ermüdungsmessungen in vier japanischen Schulen, in I. M. S. H., 1: 1905, 53–100.

(32) Schmey, Ueber Modificationen der Tastempfindung, in DuBois-Reymond's Archiv. (Physiol.), 1884, 309–312.

(33) R. Schulze, Aus der Werkstatt d. exp. Psych. u. Pädagogik, Leipzig, 1909. Pp. 292.

(34) M.-C. Schuyten, Sur les méthodes de mensuration de la fatigue des écoliers, in Ar. P., 2: 1903, 321-6.

(35) M.-C. Schuyten, Comment doit-on mesurer la fatigue des écoliers, in Ar. P., 4: 1904, 113–128.

(36) T. Simon, L'interprétation des sensations tactiles chez les enfants arrières, in A. P., 7: 1900 (1901), 536–558.

(37) L. Solomons, Discrimination in cutaneous sensations, in P. R., 4: 1897, 246–250.

(38) R. Sommer, Ausstellung v. exp. psych. App. u. Methoden, 1904.

(39) G. Tawney, The perception of two points not the space-threshold, in P. R., 2: 1895, 585–593.

(40) G. Tawney, Ueber die Wahrnehmung zweier Punkte mittelst des Tastsinnes, mit Rücksicht auf die Frage der Uebung, in Ph. S., 13: 1897, 163–222.

(41) Helen B. Thompson, The mental traits of sex, Chicago, 1903. Pp. 188.

(42) J. van Biervliet, L'asymétrie sensorielle, in Bull. Acad. Royale des Sciences, etc. de Belgique, 34: Série 3, 1897, 326–366.

(43) J. van Biervliet, La mesure de l'intelligence, in Jour. de psych., 1: 1904, 225-235.

(44) T. Vannod, (*a*) La fatigue intellectuelle et son influence sur la sensibilité cutanée, in Rev. Méd. de la Suisse Romande, 17: 1897, 21. (*b*) La méthode esthésiométrique pour la mensuration de la fatigue intellectuelle, in Rept. 1st Intern. Cong. on School Hygiene, Nürnberg, 1904, vol. ii.

(45) K. Vierordt, Ueber die Ursache verschiedenen Entwickelung des Ortssinnes der Haut, in Pflüger's Archiv, 2: 1869, 298–306. Also Die Abhängigkeit der Ausbildung des Raumsinnes der Haut v.d. Beweglichkeit der Körperteile, in Zeits. f. Biol., 6: 1870.

(46) A. Volkmann, Ueber den Einfluss der Uebung auf das Erkennen räumlicher Distanzen, in Ber. d. Sächs.-Ges. d. Wiss., 10: 1858, math. phys. Abth., 38–69.

(47) L. Wagner, Unterricht u. Ermüdungsmessungen an Schülern des neuen Gymnasiums in Darmstadt, in S. Z., 1: 1898, Hft. 4, Pp. 134.

(48) Margaret Washburn, Some apparatus for cûtaneous stimulation (with cut), in A. J. P., 6: 1894, 422–3.

(49) Margaret Washburn, Ueber den Einfluss der Gesichtsassociationen auf die Raumwahrnehmungen der Haut, in Ph. S., 11: 1895, 190–225.

(50) E. H. Weber, De pulsu, resorptione, auditu et tactu, Leipzig, 1834.

(51) C. Wissler, The correlation of mental and physical tests, in P. R. M. S., 3: No. 6, 1901. Pp. 62.

CHAPTER VII

Tests of Attention and Perception

The tests included in this chapter are those commonly assumed to measure such capacities as "power of observation," "quickness of perception," "range of attention," "mental grasp," etc. They are practically confined to the sphere of visual perception, and imply that this perception takes place under active attention: thus, in general, they seek to determine the subject's capacity to perceive visual symbols when the conditions of perception are limited by short temporal persistence of the stimulus, or by other difficulties or complications that are intentionally introduced.

It has been said that experimental psychology discovered attention. Whether this be strictly true or not, every psychological experiment of necessity takes account of attention. And so, in every mental test that presupposes effort or concentration, we measure the capacity under investigation, always as conditioned by the particular degree of attention manifested at the time. It follows that a fundamental presuppositior for the comparison of the results of such tests is that they shall all be secured under the same condition of attention. In practise, we find that the best way to secure this constant degree of attention is always to exact the maximal degree. Yet, in so far as the capacity to attend does differ in different individuals and in the same individual at different times, just so far our tests of various other capacities, such as discrimination, retention, and the like, are often felt to be measures of attention, quite as much as measures of these other capacities.

Despite this fact, or perhaps on account of it, a direct measure or test of degree of attention is difficult to secure. In theory, since attention is a condition of consciousness in which certain constituent processes are clear and prominent, attention is directly measurable in terms of clearness. In practise, we must, in all probability, content ourselves with an attempt to measure attention

indirectly, not by any single test, but by a series of tests, all of which exact maximal effort: even then, it must be recognized that we measure, not the process or condition of attention itself, but a product or concomitant of that condition.[1]

A fundamental source of difficulty in these tests of attention lies in the fact that, with repetition, or even with the progress of the first trial, the task assigned becomes progressively easier; a tendency toward automatism appears, and the tax on attention diminishes proportionately.

TEST 24

Range of visual attention.—In a single 'pulse' of attention only a small number of impressions can stand out clearly: the area or span of consciousness is definitely limited. In the sphere of vision, we find that if we give but a single glance at any heterogeneous collection of objects, such as the goods displayed in a store-window, or the jumble of odds and ends in an old tool-chest, we are able to grasp and enumerate only a very few, perhaps four or five, of these objects.

For the scientific study of the area or range of visual attention, psychologists employ some form of short exposure apparatus, or tachistoscope[2] (Greek, *tachistos*, very rapid, and *skopein*, to view). The essential idea of a tachistoscope is to furnish a field upon which *S* may fixate his glance and attention, and to supplant this field for a brief instant by another which contains the test-material. There is, then, a pre-exposure field (which contains a fixation-mark), an exposure-field, and a post-exposure field. The contents of the exposure-field depend, of course, upon the object of the experiment. In the main, the tachistoscope has been most used for the experimental investigation of the process of reading, and, accordingly, with an exposure field containing printed texts, isolated words, nonsense syllables, single letters, etc., but it has also been used for determining the range of attention for the visual apprehension of groups of lines, geometrical drawings, objects, colors, etc.

[1] The recent work of Geissler (The measurement of attention, in A. J. P., 20: 1909, 473–529) has shown that the theoretical conclusion of p. 221 is justified in actual experimentation, though his methods are too elaborate to serve the purpose of simple functional tests.

[2] The name *Tachistoskop* was first employed by Volkmann (27)

As Dodge has remarked, "no psychological instrument is subject to greater modification in response to special experimental conditions than exposure apparatus," and it may be added that in no other experiment are the results more evidently conditioned by the form of apparatus and type of procedure employed. For these reasons, it is advisable to review briefly the development of the instrument in the light of the experimental requirements.

Wundt has formulated the essentials of a good tachistoscope as follows (29):

(1) The exposure must be short enough to preclude eye-movements.

(2) The arrangement of the fixation-mark and of the stimulus must be such that all the constituents of the exposed object can be seen with at least approximately equal distinctness, *i.e.*, the exposure-field must coincide with the ocular field of direct vision.

(3) The exposure of all parts of the field should be simultaneous, or so nearly so that there shall be no noticeable time-differences in the illumination of the various regions.

(4) Retinal adaptation must be favorable, and sudden transitions from dark to light must be avoided.

(5) Persistent after-images must be avoided.

(6) The duration of the retinal excitation must be limited enough to preclude roving of attention over the exposure-field.

(7) A ready-signal must be given at an appropriate time before the exposure.

Further requirements given by Dodge[1] are as follows:

(8) The relative illumination of the pre-exposure, exposure, and post-exposure fields should be capable of experimental modification.

(9) The exposure should be noiseless and free from distraction.

(10) It should be possible to arrange for monocular or binocular observation.

The earliest exposure apparatus, as used by Dove, Zöllner, and Helmholtz, and described by Helmholtz (14, p. 710), made use of the electric spark for illumination. Here, although the duration of the spark is but 0.00004 sec., the retinal excitation of the darkness-adapted eye is longer than is obtained from an exposure of the same object for 0.01 sec. in daylight.

The use of a rotating disc for tachistoscopic experimentation is illustrated by early devices described by Helmholtz (p. 514), and more fully by Exner (10), and employed, though for somewhat different purposes, by Exner and by Baxt (1). An elaboration of the rotating disc apparatus was used by Goldscheider and Müller (11), and the principle is embodied in a recent rotating mirror tachistoscope by Wirth (29). Here the fixation-field

[1] See also Erdmann and Dodge (8, pp. 94 ff.) for special requirements for the investigation of reading.

is a virtual image obtained from the revolving mirror and the exposure-field is a real image viewed through an adjustable slit in the circular mirror, which is revolved by a constant speed motor. A very simple adaptation of the rotating disc apparatus is illustrated in the weight-driven sectors used by Quantz (24) and another by the apparatus prescribed by Titchener (26, ii., p. 201), while the instrument here employed for the range test may be regarded as a modification of this form.

A well-known exposure apparatus is the fall or gravity tachistoscope employed by Cattell (4), Huey (17, 18) and others. In this instrument, exposure is accomplished by the drop of a guillotine-like screen, perforated with a horizontal slit of variable width, before a card bearing the exposure-field. To render the exposure more nearly simultaneous, to afford a wider range of exposure-times, and to secure more accurate fixation, this instrument is now made with taller columns, with an Atwood-machine attachment and with a new form of fixation-field, as illustrated by Zeitler (31, p. 381). Monocular vision with the aid of a reading telescope is also introduced. A large demonstration form of the fall-tachistoscope is figured by Wundt.

A pendulum exposure apparatus is shown in Wundt (p. 400), also in Zimmerman's catalog (1897, p. 8). The pendulum has also formed a constituent part of other tachistoscopes, *e.g.*, as a device for interrupting a flash of light, as in Sanford's dark-box for testing legibility of various forms of type (25), or in Dodge's tachistoscope.

The horizontal exposure used by Volkmann is seen again in Hylan's simple rubber-band and shutter type of tachistoscope (19, pp. 395 and 509).

Another simple device is the superposition of a photographic shutter 4 cm. in diameter over the test-material (Binet, 3).

The use of a transparent mirror gives the cue to the construction of Dodge's latest instrument (6, 7, also described in Judd, 20*b*, p. 234), which embodies all the requirements above cited. This instrument is to be recommended for those who wish to do careful work under experimentally varied conditions. Its only disadvantage is the necessity of employing a high-power illuminant, such as the arc light or 150 C.P. stereopticon incandescent lamp. The instrument costs about $20 without lamp, discs, exposure pendulum and other accessories.

The apparatus used by Erdmann and Dodge (10, 98 ff.) and by Becher (2) is of the camera obscura type, in which an image of the exposure material is cast by a beam of light upon a ground glass field, and the illuminating beam interrupted by suitable devices.

The controversies concerning the interpretation of the results secured by these various instruments have to deal, so far as technique is concerned, with four main problems:—(*a*) How essential is absolute simultaneity of exposure? (*b*) How carefully are convergence and accommodation controlled by the fixation-point? (*c*) How long is, and how long should be, the actual

duration of the retinal excitation set up by the exposure? (*d*) What are the optimal conditions of general and local adaptation? A brief discussion of these issues is essential to the intelligent examination of the results of tachistoscopy.

(*a*) *Simultaneity of exposure.* Erdmann and Dodge contend that absolute simultaneity of exposure over the entire field is a prerequisite for successful tachistoscopy, and they therefore discount experiments, especially those of Cattell, that have been performed by falling screen and disc tachistoscopes. Wundt, however, maintains that Cattell's fall-tachistoscope gives us virtual simultaneity when we regard, not the physical exposure, but the retinal excitation which it induces. Hylan's demonstration that a series of letters may be exposed seriatim from right to left or from left to right indifferently, so far as the resulting experience is concerned, appears to confirm Wundt's position.

(*b*) *Fixation.* In the Cattell instrument, the fixation-point lies 3 mm. in front of the test-card. Sanford pointed out that this arrangement produced faulty accommodation with a tendency to double-images. Erdmann and Dodge show that at the ordinary reading-distance,[1] the displacement amounts to 0.6 mm., and this fact affords, in their opinion, the explanation of the relatively small number of words that could be read by Cattell's *S*'s in comparison with their own. As already noted, this difficulty is remedied in the improved fall-tachistoscope. In the simpler moving-screen instruments in which the fixation-point is placed on the screen, there is undoubtedly a tendency toward ocular reaction, *i.e.*, toward following the movement of the fixation-point, with consequent disturbance of fixation for the ensuing exposure.

It should be made clear that the center of attention does not necessarily coincide with the fixation-point. Furthermore, there is, strictly speaking, no such thing as a fixation-point, and no "punctiform functional center of the retina" (Dodge, 7) on which impressions are centered when attended to. The eye-muscles of a perfectly normal *S* are subject to relatively slow fluctuations of tension (fixation pseudo-nystagmus), so that what is termed the fixation *point* is really a fixation *area*. There are also slight movements of the head, due to pulse, breathing, fluctuating muscular tonicity, etc., even when elaborate forms of head-rest are provided.

(*c*) *Duration of exposure.* The times chosen for objective exposure have varied from 1 to 1000σ (σ representing 0.001 sec.). Thus, Cattell used from 1σ up, and placed the optimal time at 10σ. Messmer concluded that 2σ was sufficient after practise. Goldscheider and Müller employed 10σ, Huey 15σ, Zeitler 10–20σ, Binet about 70σ, Hylan 3.6σ in seriatim letter exposures

[1] No careful study of the optimal reading-distance seems to have been made. From 30 to 40 cm. is ordinarily used, but Hylan performed tests with a distance of 1 m.

and 42σ for other tests, Erdmann and Dodge 100σ. The primary object in using the smallest of these exposures has been to eliminate the possibility of eye-movement, and in the case of some investigators, *e.g.*, Zeitler, to eliminate the roving of attention, and to simplify, if possible, the psychical processes subsequent to exposure.

In so far as eye-movement is concerned, it is scarcely necessary to reduce the exposure-time to the limits used by most investigators, since recent studies of eye-movements and of reading-pauses indicate that the average eye-movement reaction is above 150σ (Dodge, 7), while the average duration of normal reading pauses is placed at 185σ (Huey), with a lower limit of 100σ, or, in the case of forced rapid reading of completely familiar texts, of not less than 70–100σ (Dodge).

In so far as roving of attention and other psychological complications are concerned, it seems probable, as will be shown later, that there is no adequate sanction for the use of these minimal exposures, and Dodge is doubtless correct in saying that an exposure of 100σ "guarantees what is physiologically a single visual act."

(*d*) *Adaptation.* This consideration of objective exposure-times is, however, but introductory to the real problem. In brief, the objective duration of exposure is not the subjective duration of exposure. At the moment of exposure there is a slight lag in excitation, due to retinal inertia, but this is a negligible matter of some 2σ. When, however, the objective exposure ceases, the retinal excitation persists for a relatively long time. As the duration of this persistence of excitation depends very largely upon the adaptation of the retina and the relative brightness of pre-exposure, exposure and post-exposure fields, it is evident that the published exposure-times are meaningless unless we know the conditions of general and local adaptation under which they were made. Thus, an exposure of 30σ, with all three fields of equal illumination, is estimated to be approximately equal to an exposure of 1σ with a pre- and post-exposure field of black. In further illustration, Wundt's controversy with Erdmann and Dodge (Wundt, Erdmann and Dodge, 9, Becher) hinges in part upon the question:—how long was the retinal excitation (*Bildzeit*) that was secured under the experimental conditions used by Erdmann and Dodge when their objective exposure was 100σ? Wundt estimates 250σ, Erdmann and Dodge, 150σ, Becher, less than 150σ. Into the merits of these estimates, it is not here the place to go. In general, it is well to work in diffuse day-light adaptation. As regards local adaptation (relative brightness of the three fields directly used in the test), we may conclude that it is well not to have the post-exposure field stimulating, *i.e.*, brighter than the exposure-field, and that a black pre- and post-exposure field affords the most brilliant exposures, with longest retinal effects. Under these conditions, which are supplied by the apparatus used in this test, an exposure of 50–75σ will give a well "cleared-up" image (to employ Dodge's term) and will eliminate eye-movement.

APPARATUS.—Disc tachistoscope (Fig. 49).[1] Frosted tubular lamp, 16 C. P. Two 4-inch clamps. Blanks of cardboard 9 cm. square. Two complete sets of Willson's gummed black letters and

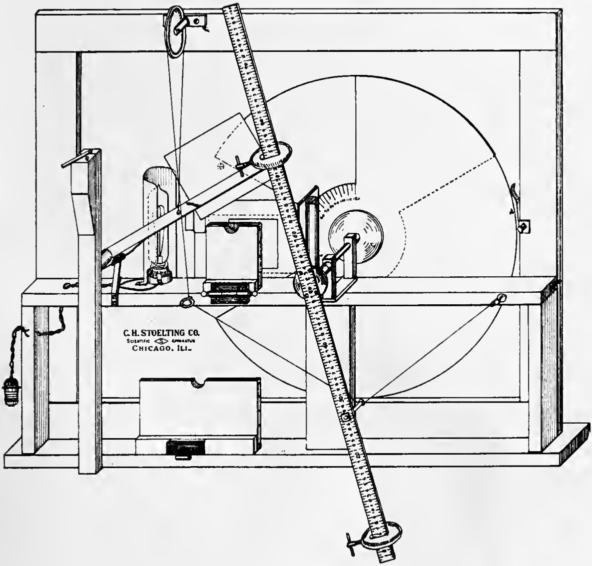


FIG. 49. DISC TACHISTOSCOPE.

figures, Size 3. Drawing ink and ruling pen. Head-rest,[2] with suitable supports and clamps. [Fifty-vibrations dry-contact fork (Fig. 50.) Dry battery. Connecting wire.]

[1] This instrument, devised by the author and employed by him in studying the effect of practise on the range of attention does not fulfill all of the requirements of the ideal tachistoscope, but it has the merit of being relatively inexpensive, simple in operation and construction, and of answering satisfactorily for comparative tests. It has also been designed with a view for use both for these short and for much longer exposures (Test 25).

[2] Suggestions for the construction of head-rests are given by Judd.

PRELIMINARIES.—Prepare a series of exposure-cards by use of the gummed letters and figures. Paste the letters smoothly and evenly, and center the series on each card. The test-object should include isolated letters, groups of letters, or letters mixed with digits in nonsense arrangement, in numbers from two to eight or ten per card. Prepare other cards with short words, short sentences, or columns of digits. Prepare still others with the aid of drawing ink so as to present regular or irregular series of lines, geometrical fig-

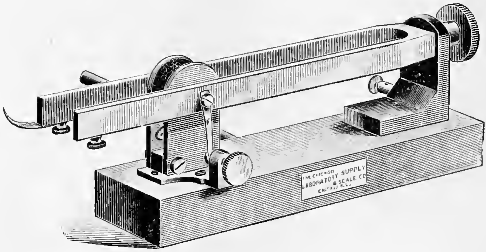


FIG. 50. FIFTY-VIBRATIONS DRY-CONTACT FORK.

ures, surfaces, etc., in varied fashion.[1] It is well to confine the objects, in so far as possible, to an area 50 mm. wide and 35 mm. high. Exposure cards can also be formed, after the method used by Huey, by pasting bits of printed texts of various dimensions on the blanks, or colored surfaces may be introduced *ad libitum*.

Set up the tachistoscope, and clamp or screw it firmly to the table. Connect the electric light to a circuit of proper voltage. By means of the supports and clamps, adjust the head-rest so that *S*'s eyes shall be about 40 cm. in front of, and slightly higher than, the exposure card.

Adjust the tachistoscope for a point-exposure of 60σ.[2]

[1] For suggestions, see Huey, 18, pp. 75 ff., and Binet, 3, p. 349.

[2] The time of total exposure of several rows of objects is not the same as that of the exposure of any single point. If t is the time required for the notched portion of the disc to pass a given point on the exposure field, and L is the time of exposure of the group of objects to be determined, and if H and h are the heights of the total exposure field and of the portion of the field occupied by the objects, respectively, then $L = t\,(1+\frac{h}{H})$.

The pendulum-arm is attached with its mm. scale in view at the back of the instrument, and with the zero end of its scale lying upon the release-lever when the instrument is ready for exposure. To secure an exposure of 60σ, *E* should set the first weight at 5 cm. (measured at the edge of the weight nearer the lever), should set the second weight at 60 cm. and should open the sector 25 deg. He may fasten the overlapping edges of sector and disc by pushing a small paper-clip over them at the periphery. The entire disc is to be tightened up upon its axis in such a position as to bring the square opening in the disc in alignment with the square opening in the screen.[1]

See that the device that lifts the fixation-card works smoothly and quietly. The movement of this card should be invisible to *S*, and the stimulus-object should be entirely unobstructed by it at the moment of exposure

METHOD.—Seat *S* so that his head is supported in the head-rest without undue strain (Fig. 51). Set the tachistoscope for an exposure, and place a very simple exposure-card in the holder. In this position of the disc, the exposure-card is hidden by the fixation-card, which is visible through an opening in the disc. Instruct *S* to fixate the cross of the fixation-card as accurately and intently as possible, and, when he feels that his attention is thoroughly prepared, to signal for the release of the disc.[2] Let him

[1] If he wishes to measure the time of exposure accurately, *E* may attach a piece of smoked paper temporarily to the back face of the disc, connect the 50-vibs. fork with the battery, and adjust the fork so that its recording point (or the recorder of a sensitive signal-magnet, *e.g.*, a Deprez signal with which the fork is electrically connected) leaves its curve traced upon the paper when the disc is released. He may then determine the time of exposure by counting the number of 'waves' recorded while the notched portion of the disc is passing the center of the exposure field.

Or he may apply the fork directly to a smoked kymograph drum (see Test 10) and record the duration of the exposure with a signal-magnet on a parallel tracing. For this purpose, the two clips fitted with light connecting wire, with which the instrument is provided, are placed on the periphery of the disc in such a manner that, at the moment of exposure and at the moment of occlusion of the center of the field, they make electrical contact with the copper brush which is fastened to the frame of the instrument.

[2] The rather unusual procedure of allowing *S* to control initiation of the experiment is justified here by the fact that the signal for the release may be very simple, and that we are more likely to secure maximal attention from *S*. If, after several practise trials, it is evident that *S* is distracted by this procedure, *E* should revert to the usual method of giving the ready-signal himself. The practise series, in any event, need be no longer than is required to accustom *S* to the general setting of the test.

then report, orally or preferably by drawing, what he has seen. Unless attention is manifestly poor, do not repeat the exposure.[1]

Introduce more and more complex cards of the type under investigation until a limit is reached beyond which *S* cannot carry his observation. For the experiment proper, use in the

FIG. 51. TESTING THE RANGE OF ATTENTION.

main simple series of consonants, and drawings, but unless there is some special reason for it, do not change without notice from the one type of exposure-card to the other.

VARIATIONS OF METHOD.—Test the range of attention with other forms of material, *e.g.*, digits, nonsense syllables, words, and

[1] There is, however, good precedent for using repeated exposures, if desired: Cattell used a series of 5 exposures; Huey occasionally, and Binet regularly, made as many as 20 exposures; Titchener prescribes an indefinite number of exposures. This procedure is based on the assumption that no more is actually *seen* in 20 exposures than in one, but that the series of exposures determines the limits of assimilative capacity. This method seems unnecessary, especially for comparative purposes.

especially with short sentences and drawings. Test the assimilative completion of word skeletons (groups of characteristic letters, mutilated or misspelled words, as illustrated under Results, 13, *c*). Vary the time of exposure, especially toward the shorter times. Try changes in the backgrounds and fields by covering the screen and disc with white, gray and black papers in different combinations, and using gray or black blanks for exposure-cards. Try the cumulative method of exposure. Try reading distances greater than 40 cm. Try various sizes, forms, or colors of type.

TREATMENT OF RESULTS.—*S*'s rank is measured in general by the number of concrete objects correctly reproduced. This rank should be computed separately for different forms of exposure-card (if other than isolated letter tests are used) and for letter-series of different lengths, *e.g.*, for 6-place, 7-place, 8-place series, etc. In the case of letters or words, the simplest method of ranking is to assign one unit for each letter or word correctly reproduced, but to deduct 0.5 for errors of insertion or transposition. Thus, if the stimulus-card was FRMUTH, then FRMTUTH or FMRUTH would be ranked 5.5, whereas FRMUT or FRMUH would be ranked 5. If desired, double errors, *e.g.*, transposition and insertion, as FMRTUTH, may be doubly discounted, but it is simpler to count this rendering also as 5.5.

In the case of more complex exposure-cards, such as drawings, it is sometimes possible to rank *S* on the basis of the number of lines reproduced, discounting 0.5 for lines incorrectly placed with respect to the total figure; in other cases, it is more satisfactory to assign a subjective estimate of the general fidelity of the reproduction on some arbitrary scale, *e.g.*, 10 for a perfect reproduction, 0 for absolutely nothing, and intermediate ranks in proportion.

If the method of serial exposures is followed, *S*'s rank can often be indicated by the number of exposures necessary to obtain accurate reproduction.

RESULTS AND CONCLUSIONS.—(1) When a series of *unrelated objects* is exposed, the average number of impressions that can be grasped in a single exposure lies between four and five. While, occasionally, *S*'s may grasp as many as seven impressions, this is

usually due to a more or less well-recognized tendency to group or unify the objects in some manner. Table 30 summarizes the author's experiments (28) upon four college students with *letter-series*, 100σ exposure, and with the use of apparatus like that prescribed, save that the pre- and post-exposure fields were white. It will

TABLE 30

Average Number of Letters Read Correctly in one Exposure (Whipple)

OBSERVERS	FIVE-PLACE SERIES	SIX-PLACE SERIES	SEVEN-PLACE SERIES	ALL
Mr. B.	4.85	5.09	5.25	5.06
Mr. E.	4.84	4.49	4.48	4.53
Mr. N.	4.74	4.92	5.38	4.97
Mr. T.	4.51	4.86	4.40	4.71
Average	4.74	4.84	4.88	4.82

be seen that the average performance for different *S*'s under different conditions is quite similar, and that it lies between four and five impressions. The 6- and 7-place series are somewhat more favorable because the *S*'s can occasionally reproduce 6 or 7 impressions.

These results are in close accord with those of other experimenters.[1] Cattell's tests (4) place the average limit for digits at 5, for letters at 3-4, and more often 3, though one *S* could grasp 6 letters. Erdmann and Dodge found that 6-7 letters could be read at times. Zeitler (30) points out that, while a series of consonants has an assimilative limit of 4-7, one may grasp 5-8 impressions, if vowels are interspersed. In the author's experiments, it was found, similarly, that the 7-place series that were read successfully were almost always those containing vowels which permitted the formation of nonsense syllables, *e.g.*, WAEGZME, KMDEMBH.

[1] The horizontal seriatim exposure used by Hylan is so little comparable with the whole-field exposure of others that his results may not be expected to conform with theirs. Hylan found that, with an exposure of 6 letters at the rate of 3.6σ per letter, an average of 1.9 letters was read, while a longer exposure, 42σ for the entire card, permitted 2.6 letters to be read.

The statement that the first and last letters are those usually seen clearly is not so easily confirmed and does not apply to the results of all *S*'s.

(2) *S*'s not infrequently report that more is seen than can be remembered a moment later when the report is given. In such cases, it is still often possible to state whether or not a given character was present.

(3) Despite the meaningless character of isolated letters, a series once exposed may be sufficiently well remembered to be recognized if used in the experiment again, even after a lapse of several days.

(4) When familiar syllables are combined to form *nonsense words, e.g., lencurbilber,* 6-10 letters can be grasped in one exposure (Zeitler).

(5) In the reading of *isolated words, i.e.,* collocations of words that do not make sense, the results of different investigators show a lack of accordance. As in the case of letters, Cattell's results indicate a lower range than those of other investigators. Cattell placed the limit of grasp at 2-3 short one-syllabled words. Erdmann and Dodge (8) found that, in a single exposure, 4 isolated words can almost always be read, and very often 5. Again, single words of 19-22 letters can be read in one exposure without roving of attention. Some of Zeitler's *S*'s could even read a 25-letter word, such as *Aufmerksamkeitsschwankung.* Becher read 26-letter words with a single, electric spark exposure. In general, the difficulty of grasping words does not increase in proportion to their length, while Cattell's reaction experiments showed that short words can be named more quickly than letters.

(6) When short words are combined into *simple sentences,* it is found that the average reading capacity is 4-6 words. Zeitler's *S*'s read such sentences when the total number of letters was 20-30; Erdmann and Dodge report 4-6 words of 2-10 letters each; Cattell found the average amount 4 words, occasionally 6, though *S* could read at times a 7-word sentence. One of the longest correctly read sentences contained 34 letters: "*Eine Tochter muss ihrem Vater gehorchen.*"

Huey (17, 18) exposed for 15 σ lines cut from magazines. His results show clearly the striking individual differences in the read-

ing range, which practically every investigator has noted;[1] thus, one *S* read on the average continuously and correctly 10.25 mm. of the line, another 21.33 mm., a third 23.80 mm. and a fourth 32.40 mm.[2] Under very favorable conditions (attention, subject-matter, etc.), one of Huey's *S*'s could read a stretch of line 50 mm. long (about half a line), *e.g.*, such phrases as "condition of consciousness," or "the whole body converges," but these are not ordinary performances.

In exposures of this sort the amount read to the left and to the right of the fixation-point is not at all equal, but varies in either direction according to the subject-matter, *e.g.*, in the second phrase above, the fixation-point was at the *o* of *whole*. The tendency is, as might be anticipated, to read more to the right than to the left. "In almost every case in which a large amount is read, far more is read to the right of the fixation-point than to the left."

In these, as in other exposures, the extent of reading is curtailed in proportion as the word-groups resemble isolated words, as when divided by punctuation-marks.

There is evidence to show that the unit of reading is the word, in as much as *S*'s see, or at least report, words or phrases rather than letters, even at the ends of the sections read.

(7) If the exposure-card is composed of a series of 4-15 ruled perpendicular *lines*, 2 mm. apart, Cattell found that *S*'s could give the correct number of lines exposed, up to 4-6 only. Similarly, Goldscheider and Müller (11) tried various combinations of straight and curved lines (10 σ exposure), with the result that, if the arrangement was quite irregular, only 4-5 constituents could be grasped, but in proportion as the arrangement became more symmetrical (thus facilitating grouping or unitizing), the number of constituents that could be grasped was increased. A symmetrical arrangement of simple perpendicular strokes increased the observation-limit to 7, while a combination of straight lines into squares, symmetrically arranged, permitted the apprehension of 5 squares, and hence of 20 constituent lines. Similarly, experiments with semicircles, ellipses, etc., confirmed the general principle that the num-

[1] Cattell concludes that there is a decided difference in the sensitiveness of the retinas in different *S*'s, but it is quite as likely that these individual differences are as much central as peripheral.

[2] In the type used, 40 mm. was equal approximately to 26 letter-spaces

ber of constituent elements grasped in a short exposure is a function of the degree of combination which these elements permit.[1]

(8) The apprehension of *simple geometrical forms, e.g.*, circles, diamonds, oblongs, etc., cut from black paper and pasted on a white background, is easier than that of letters. Hylan's results, compared with his results for letters above, may serve for illustration. Six *S*'s averaged 7 forms, his poorest *S* averaged 4.5, and his best *S*, 9.5 forms.[2] Quantz's test of requiring *S*'s to name aloud, in order and as rapidly as possible, a series of geometrical forms, colors or words, during an exposure of 0.5 sec. or 1.0 sec. is not strictly comparable to the short exposure tests: he found that forms could be named less rapidly than colors or words, and that, so far as forms are concerned, as many can be named with 0.5 sec. as with 1.0 sec. exposure (2.75 and 2.8, respectively).[3]

If *complex drawings* which are not clearly related to well-known geometrical figures are used, the test becomes more difficult because the visual image cannot be identified or held by the assistance of verbal associates (Binet, 3).

(9) *Practise* has a curiously small effect upon the range of attention, when once the period of preliminary habituation to the arrangement of apparatus and method is passed (Cattell, 4; Hylan, 19, p. 396). The chief feature of whatever practise can be detected is an increase in ability to group isolated impressions into combinations. "Practise tends to unite into a closer perceptive unity impressions first combined with difficulty" (Hylan). The practise effect for isolated letter series for a period of seven to ten days as found in the author's tests (28) is indicated in Table 31.[4] It

[1] An obvious illustration of this principle is seen in the reading of letters and digits themselves. For an account of Goldscheider and Müller's tests of the constituents of digits, etc., consult Huey (18, pp. 78 f.).

[2] Some idea of the qualitative factors which influence the perception of liminal visual forms may be obtained from the experiments of Miss Hempstead, though these were conducted with different apparatus and by a different method.

[3] Huey's statement (18, p. 54) that, according to Quantz, more could be read in a short than in a long exposure, is not substantiated by Quantz's tables, which merely show that *relatively* more can be read in a half-second than in a second exposure.

[4] The four *S*'s did not take the same series, nor work for the same length of time, so that the data are too few to permit accurate averaging into three periods, save for *N*., and for *E*. in the 7-place series.

is evident that if we discount the improvement due to adaptation, there is but a small enlargement of the range through practise.

TABLE 31

Effect of Practise upon the Perception of Letters (Whipple)

OBSERVER	FIVE-PLACE SERIES			SIX-PLACE SERIES				SEVEN-PLACE SERIES		
	B.	N.	T.	B.	E.	N.	T.	E.	N.	T.
First period.................	4.87	4.44	4.50	5.03	4.75	4.38	4.73	4.25	4.90	3.83
Later period.................	—	4.87	—	—	—	4.85	—	4.02	5.54	—
Last period.................	4.78	4.77	4.50	5.25	5.08	5.06	5.08	4.90	5.40	5.80

(10) The relation of the visual range of attention to *age* has been studied carefully only by Griffing (12), but by a method so peculiar[1] as to make the applicability of his results to ordinary conditions rather dubious. Table 32 presents Griffing's results in terms of the total number of letters correctly read in a series of 10 exposures of 6 letters each. Griffing concludes from these data that the number of visual impressions perceived "is a function of individual growth, reaching its maximum only when the observer is fully developed."

TABLE 32

Relation of Visual Range of Attention to Age (Griffing)

NUMBER OF S'S	AGE	LETTERS READ
39	7–9	4
77	10–12	13
73	13–15	18
132	16–18 +	27

[1] Griffing's method was to expose with a fall-tachistoscope for 0.1 sec. or 1.0 sec., to a group of 10 to 20 *S*'s, six printed capital letters, 48 mm. high, but to vary the interval between the ready-signal and the exposure from 6 sec. to 1.5 min., or even 4 min., without the knowledge of the *S*'s, with the idea of testing "the observer's powers of prolonged attention" by keeping them waiting an indefinite time for the exposure. In view of the well-known irregularity and fluctuation of attention under such conditions, it seems clear that the degree of attention present at the moment of exposure, at least in a short series of tests, is almost a matter of chance.

(11) Griffing found no difference between the *sexes*.

(12) There is only questionable evidence of a relation between the *range of visual attention and mental ability*. Griffing divided his pupils into three groups on the basis of teachers' estimates, and found that his 'A' group had a somewhat higher average range of attention than the other groups, but that there were marked exceptions, so that "many pupils must have good powers of attention even when they show no evidence of them to their teachers." It is possible, however, that the outcome of Griffing's test is more dependent on the interest and good-will of the pupils than on their intelligence, for, as Griffing himself says: "Children of the most active minds would be most interested in novel experiences." Cattell states that in his tests "obtuse porters" required three times as long as educated persons to read a letter or word. Binet could not differentiate his bright from his dull children by the exposure of single words, but could differentiate them very clearly by the exposure of a drawing 20 times in succession.[1]

(13) The *qualitative analysis* of the perceptive processes concerned in reading during short exposures has developed differences of opinion with regard to the following points: (*a*) Is it possible for roving of attention to occur in exposures which are sufficiently short to eliminate eye-movement? (*b*) Do we apprehend words by wholes or by parts? (*c*) Are there certain letters or combinations of letters which give the cue for the perception of words, and if so, what are these letters? What share does the total length and general visual contour of a printed word have in its perception? (*d*) Do different readers adopt different methods of reading? Current opinions upon these obviously interrelated problems may be summarized as follows:

(*a*) *Roving versus distributed attention.* The question here is: is attention so distributed during the retinal excitation that we apprehend simultaneously all phases of what is read, or does attention 'rove' from one portion

[1] Binet's *S*'s drew what they could of the drawing after each exposure of 70σ. An interval of 5–10 sec. was interposed between the exposure and the drawing. Only three of 11 *S*'s were able to give a correct copy within the limit set to the test; these three *S*'s were all of the group selected for superior intelligence. For reproductions of the actual drawings, consult Binet, 3, pp. 351–360. The original drawing is the left-hand one shown in the Binet-Simon graded tests, in Chapter 13, Fig. 59.

to another of the visual stimulus? The discrepancy noted between the results of Cattell and of Erdmann and Dodge (5, above) is attributed by Wundt to the use of too long an exposure with consequent roving of attention (and also, in part, to familiarity with the words to be used). Erdmann and Dodge, however, deny that roving took place during the visual phase of the experience, and assert that whatever roving occurs is confined to the motor-acoustic interpretation of what is seen: they find an explanation for Cattell's results in imperfect accommodation. They also point out that *S*'s can often recognize a word under conditions such that its constituent letters cannot be recognized, so that a rough visual image of the whole word must suffice for its recognition by associative supplementing—a fact that shows how difficult it is to distinguish between what is actually seen and what is 'centrally' supplied in tachistoscopic experiments. The varied experiments of Becher confirm the position of Erdmann and Dodge.

Hylan (19) and Pillsbury (23, ch.v.) contend that there is, strictly speaking, no simultaneous distribution of attention at all in tachistoscopy, only successive acts of attention applied to the "mental after-image." If we see four or five objects it is because "the result of a single glance persists long enough for four or five acts of attention to take place." This mental after-image is distinct from the ocular after-image, though a direct product of stimulation and conditioned by physiological processes. Hylan estimates its average duration at 2.8 sec., and believes that individual physiological differences in this duration determine the number of objects that can be apprehended in one so-called pulse of attention.

(*b*) *Reading by wholes or by parts.* Intimately connected with the question of the roving of attention is the question: do we read by wholes or by parts? It has been a time-honored idea, fortified by certain observations by psychiatrists (Grashey and Wernicke, especially) upon aphasic patients, that we read letter by letter. Cattell was convinced, however, that words, and even short sentences, if read correctly, or even if read with some false interpretation, are grasped as a whole, so that their component parts appear plainly and simultaneously in consciousness. Erdmann and Dodge similarly believe that we commonly perceive words as wholes. Goldscheider and Müller conclude that words are recognized either as syntheses of their perceived component letters or as individual units, the latter being the more usual method. Messmer reaches similar conclusions. Zeitler distinguishes between apperception (an immediate process conditioned largely by the objective image) and assimilation (a mediate process taking more time), and has sought to eliminate assimilation by using a very short exposure. He concludes that, on account of long practise and familiarity with words, we *seem* to read them as wholes, but that, in reality, there is a quick succession of processes, by a passage of attention, not from letter to letter, but from one to another of a series of prominent letters and letter-combinations. He believes that Cattell failed to note this succession because his exposures were too short, and that Erdmann and Dodge were

misled by the illusion of simultaneous assimilation. Dodge (7), however, asserts, as already indicated, that this alleged successive apprehension is really a successive review of the memory-image of an inadequately 'cleared-up' exposure. If the exposure is adequate for full clearing-up, we find simultaneous visual apprehension. Becher (2) has shown that geometrical-optical illusions, which Wundt instances as typical assimilations, can be fully apprehended in a 10σ exposure: he argues from this that Zeitler could not exclude assimilation by his short exposures.

(*c*) *Dominating and characteristic letters.* Investigators like Goldscheider and Müller, Zeitler, and Messmer, who argue against reading by wholes, do not assert that we read letter by letter, but rather that certain prominent letters which 'stand out' in the visual image give us the cue for the acoustic-motor supplementing which forms such a prominent part of the process of reading; the interpretation is also, of course, greatly assisted by knowledge of the context and of the immediately preceding words and phrases. These prominent letters are termed "dominating" letters by Zeitler, "determining" letters by Goldscheider and Müller (11); the other letters are termed "indifferent." As dominating letters, Zeitler mentions those which project above and below the line, and *x* and *z*. Messmer maintains that letters that project below the line are not so important; "they possess optically the value of small letters," and are often mistaken for these: again, they are relatively infrequent; *e.g.*, in a sample passage, Messmer found 238 letters projecting above the line to 32 projecting below.

There are also, according to Zeitler, dominating syllable-complexes, and in very familiar sentences, dominating words. The perception of these dominating elements suffices for the perception of the entire word or sentence, while indifferent letters or syllables may be altered or eliminated without affecting the reading. In reading, therefore, attention moves along a series of these dominating elements. At the first instant of exposure, the elements have a certain "elbow-room," and (if we may mix metaphors) "hover" about till they get "anchored" by assignment into some total word-complex in accordance with the sense of the passage. "The word-form remains indeterminate and is first established through the sense of the passage." Direct visual completion of the words takes place. Goldscheider and Müller, however, are convinced that the filling-out takes place only after the arousal of auditory imagery set off by the determining letters. In any event, the indifferent letters are so readily supplied that, when the determining letters only of a word are exposed, many *S*'s report that they have seen the entire word clearly, *e.g.*, *C ntr m* for the German *Centrum*. This visual clothing of word-skeletons is a striking example of Münsterberg's statement that "reproduced sensations under favorable conditions cannot be distinguished from sense-impressions." Similarly, a considerable amount of substitution will pass unnoticed, if restricted to indifferent letters; Zeitler's *S*'s, for instance, read *hallucination* when *hallneiuotion* was exposed. The common experiences of proof-reading supply further illustrations.

In their analysis of determining letters, Goldscheider and Müller found that the first and the last letters in each word, especially the first, are almost always determining. Similarly, by another form of experiment, Huey discovered that the first half of a word is commonly more important than the second half. These investigators, however, did not find that consonants were necessarily more important than vowels.

Messmer places less insistence upon dominating letters, at least for some readers, and directs attention rather to the total optical impression, or word-form, as a well-nigh exclusively significant factor in visual perception. In the relatively long pauses of normal reading, the dominating letters play a smaller role than in short-exposure experiments, since there is more opportunity for all the letters in the word to affect consciousness. For the same reason, the total word-form becomes more important. Messmer, for instance, contrasts the dull evenness of such a word as *zusammenreisen* with the optical 'rhythm' of such a word as *Verschiedenheiten*. So one might contrast the words *consciousness* or *amusement* with the words *orchestration* or *pedagogically*. Hamilton (13) concludes that in most cases the general characteristics of a word afford the cue for its recognition.

Becher, like Erdmann and Dodge, insists on the primary importance of the total optical contour, or word-form. In this contour, the so-called dominating letters of Zeitler play a determining role because they are characteristic and prominent letters, but they are not apprehended successively.

(*d*) *Individual differences in manner of reading.* Goldscheider and Müller believe that the relative importance of consonants and vowels as determining letters depends in part on the reader's ideational type. They also believe that the method of reading may shift from sentence to sentence, from word to word, or even within the same word, according to the general ease and familiarity of the reading-matter; in general, with unfamiliar words, the tendency is toward perception by letters; contrariwise, the more familiar the passage, the fewer visual cues are needed, and the stronger is the tendency to read by word-wholes.

Readers, in Messmer's opinion, may belong to a subjective or to an objective type: readers of the former type depend largely upon associative processes, and apperceive words in the main from the total character of their visual form; readers of the latter type make more use of the dominating elements, read less at a glance, and are less likely to err.

In conjunction with our earlier statement that "in no other experiment are the results more evidently conditioned by the form of apparatus and type of procedure employed," these final suggestions concerning fundamental individual differences in reading give us the basis for a tentative summary of the analytical results just set forth. The seeming lack of agreement between

the results of different investigators is, in short, due in part to the divergence in experimental conditions, in part of the divergence in type of the *S*'s employed. We do not always read by wholes: neither do we always read by successive apprehension. The unit of attention, of visual apprehension, in reading is a variable quantity. In normal reading, assuming equally easy subject-matter, the manner of reading will depend upon the type of the reader. Aside from differences in speed and fluency of reading, we may probably distinguish two fundamental types of readers, the subjective and the objective. The latter exhibit the following characteristics: in tachistoscopic tests, their attention is directed to the optical fixation-mark; their range is small, *e.g.*, three isolated letters or one 12-letter word, but they read accurately, are quite certain of what they do see, and seldom guess; for them there exists a distinct time-interval between the visual perception and assimilative interpretation or rise of meaning, and they seldom confuse these two phases. The subjective readers differ from the objective in every one of these points: they can read words lying in indirect vision; they have a range of five isolated letters or one 27-letter word; their attention is placed mainly on the interpretative or assimilative phase, and their reading occurs mainly by large word-wholes, or even by phrases, on the basis of relatively meager visual cues. These subjective readers are not, however, necessarily the faster.[1]

If, now, the subject-matter is difficult, the tendency for the subjective type of reader is toward the more extensive use of visual symbols, and here, doubtless, as in the case of short exposures, dominating letters or complexes become more important: it will depend upon the difficulty of the passage whether these letters play their chief rôle as prominent elements in the configuration of the word, or whether they are directly the object of attention.

If the subject-matter is very difficult, *e.g.*, the reading of beginners, or of an adult in a foreign language, especially if reading unfamiliar characters, such as Greek or Hebrew, we have an extreme case in which reading may, and often does, proceed letter by letter.

[1] For a more detailed discussion of these types, see Meumann, ii., 250 ff.

The reading of children, once the primary mechanical difficulties are mastered, is almost always of the subjective or interpretative type (Messmer), though for different reasons than in the case of the subjective reading of adults.

Whatever the type of reading or reader may be, we are never free from the presence of a mass of 'central' associates. The consequent processes of assimilation are a constant accompaniment of all normal reading. The attempt to eliminate assimilation by the use of extremely short exposures defeats its own end, and such exposures create artificial conditions, foreign to those of normal reading.

Notes.—Dr. W. McDougall of London has devised a test that is worthy of mention here because of the results that it has yielded and because of its evident similarity to the tachistoscopic exposure of geometrical figures. This test, which he calls the 'spot-pattern' test, may be said to measure *ability to apprehend topographic relationship.* A card 10 cm. square is pierced with eight or ten holes, each 1 mm. in diam. These perforations form an irregular pattern within an area 1.5 in. square, but always coincide with intersections of an imaginary system of vertical and horizontal lines $\frac{1}{4}$ in. apart. One face of the card is now covered with clear, dead-white paper[1] and the whole card is placed in an aperture of a screen in the path of a beam of light to *S*'s eye. By means of a photographic shutter, the light is flashed upon the card in groups of 5 flashes, each lasting 20 σ, and at the rate of one every 2 sec. After each series of 5 flashes, *S* is asked to make a map of the spots of light upon a bit of squared paper 1.5 in. in size, ruled to correspond to the system of lines above mentioned, *i.e.*, in $\frac{1}{4}$ in. squares. He is told the number of the spots, and that each spot must fall at the intersection of two lines. He attempts a new map after each 5 flashes, until he makes an accurate reproduction. It is of special interest to note that this test has given

[1] This arrangement makes it possible to work in daylight and without the interposition of any screens or fixation cards before the exposure-card, since the perforations remain invisible until lighted up by the beam of light passing through them. The test might, however, be adapted to the tachistoscope already described by using simple black dots in place of the perforations, and perhaps substituting a single longer exposure for the series of short exposures, the exact significance of which is not clear to the author.

very high correlations with ability as shown in general school performance.[1]

REFERENCES

(1) N. Baxt, Ueber d. Zeit welche nöthig ist, damit ein Gesichtseindruck zum Bewusstsein kommt, etc., in Pflüger's Archiv f. Physiol., 4: 1871, 325–336.

(2) E. Becher, Experimentelle u. kritische Beiträge zur Psychologie des Lesens bei kurzen Expositionzeiten, in Z. P. 36: 1904, 19–73.

(3) A. Binet, Attention et adaptation, in A. P., 6: 1899 (1900), 248–404.

(4) J. Cattell, The inertia of the eye and brain, in Brain, 8: 1886, 295–312; also as Ueber die Trägheit der Netzhaut u. des Sehcentrums, in Ph.S, 3: 1886, 94–127.

(5) R. Dodge, An improved exposure apparatus, in P. B., 4: January, 1907, 10–13.

(6) R. Dodge, The psychology of reading, in P. R., 8: 1901, 56–60.

(7) R. Dodge, An experimental study of visual fixation, in P. R. M.S., 8: No. 4, 1907, (No. 35). Pp. 95.

(8) B. Erdmann and R. Dodge, Psychologische Untersuchungen über das Lesen, Halle, 1898. Pp. 360.

(9) B. Erdmann and R. Dodge, Zur Erläuterung unserer tachistoskopischen Versuche, in Z. P., 22: 1900, 241–267.

(10) S. Exner, Ueber die zu einer Gesichtswahrnehmung nöthige Zeit, in Sitz. Ber. d. Wiener Akad., 58: (2), 1868, 601–632.

(11) A. Goldscheider and R. Müller, Zur Phys. u. Path. des Lesens, in Zeits. f. klin. Med., 23: 1893, 131–167. For critical summary, see Wallaschek in Z. P., 7: 1894, 228–231.

(12) H. Griffing, On the development of visual perception and attention, in A. J. P., 7: 1896, 227–236.

(13) F. Hamilton, The perceptual processes in reading, in Arch. of Psych., No. 9, December, 1907 (Columbia Univ. Contr., etc. 17: No. 1).

(14) H. Helmholtz, Handbuch d. physiol. Optik, Hamburg and Leipzig, 1896.

(15) Louise Hempstead, The perception of visual form, in A. J. P., 12: 1901, 185–192.

(16) E. B. Huey, Preliminary experiments in the physiology and psychology of reading, in A. J. P., 9: 1898, 575–586.

(17) E. B. Huey, The psychology and physiology of reading, in A. J. P., 11: 1900, 283–302 and 12: 1901, 292–312.

(18) E. B. Huey, The psychology and pedagogy of reading, N. Y., 1908. Pp. 445.

---

[1] C. Burt, Experimental tests of general intelligence, in B. J. P., 3: December, 1909, 94–177, especially 150 ff.

(19) J. P. Hylan, The distribution of attention, in P. R., 10: 1903, 373–403 and 498–533.

(20) C. Judd, (*a*) Laboratory manual of psychology, N. Y.,1907. Pp. 124. (*b*) Laboratory equipment for psychological experiments, N. Y., 1907. Pp. 257.

(21) O. Messmer, Zur Psychologie des Lesens bei Kindern u. Erwachsenen, in Z. g. P., 2: 1903, 190–298.

(22) E. Meumann, Vorlesungen zur Einführung in die exp. Pädagogik u. ihre psychol. Grundlagen, 2 vols., Leipzig, 1907. Pp. 555 and 446.

(23) W. B. Pillsbury, Attention, London and N. Y., 1908. Pp. 330.

(24) J. Quantz, Problems in the psychology of reading, in P. R. M. S., 2: No. 1, 1897.

(25) E. C. Sanford, The relative legibility of the small letters, in A. J.P., 1: 1888, 402–435.

(26) E. B. Titchener, Experimental psychology, vol. 1, Qualitative experiments, N. Y., 1901.

(27) A. Volkmann, Das Tachistoskop, etc., in Sitz. d. k. s. Ges. d. Wiss Leipzig (math.-phys. Cl.), 1859, 90 ff.

(28) G. M. Whipple, The effect of practise upon the range of visual attention and of visual apprehension, in J. E. P., 1: 1910, 249-262.

(29) F. Wirth, Das Spiegeltachistoskop, in Ph.S., 28: 1903, 687–700.

(30) W. Wundt, Zur Kritik tachistoskopischer Versuche, in Ph.S., 15: 1899, 287–317, and 16: 1900, 61–70. See also his Völkerpsychologie, vol. 1, Die Sprache, 1900, 530 ff., and his Grundzüge d. physiol. Psychologie, 5th ed., 1903, 3: 611 f. (Note: the first half of vol. 1 of the 5th ed. is available in English, trans. by E. B. Titchener, N. Y., 1904. Vol. 1 of the 6th German ed. appeared February, 1908.)

(31) J. Zeitler, Tachistoskopische Versuche über das Lesen, in Ph. S., 16: 1900, 380–463.

TEST 25

Visual apprehension.—This test is, in many respects, similar to that of the range of visual attention (Test 24), but it differs in certain important particulars: the period of exposure is increased from a small fraction of a second to several seconds, and the test-object is correspondingly more complex. Both eye-movement and roving of attention take place, so that we cannot speak of the range of attention;[1] we are measuring, rather, the capacity

[1] If one accepts the view of Pillsbury (4) and Hylan (3) that attention in tachistoscopy is really successive, this test differs from the former only in degree: we now give opportunity for a greater number of successive acts of attention.

to apprehend a number of disparate objects by visual examination during a short period.

Tests of this type have been variously designated as tests of "quick perception," of "observation," of "degree of attention," or even as "memory tests." The term "apprehension," however, seems suited to describe the psychological processes under examination.

Very few investigators have employed this test under experimental conditions, but it has been long used as a source of amusement in competitive parlor games, and it has been urged by some writers, particularly by Miss Aiken (1), as a source of mental training[1] in the form of systematic schoolroom exercises. Quantz (5), in his investigation of the psychology of reading, used some exposures of 2 sec.: the author (7) has made a more extended study of visual apprehension, in order to see whether the remarkable practise-effects alleged by Miss Aiken could be confirmed under laboratory conditions.

Two forms of the test are described: variations can be easily worked out to suit special conditions.

A. THREE SECONDS' EXPOSURE WITH THE TACHISTOSCOPE

APPARATUS.—Disc tachistoscope[2] and other material as in Test 24, save that the cardboard blanks are replaced by blanks of stiff

[1] The famous conjurer, Robert Houdin (2) used what he termed "perception by appreciation" as a basis for certain feats of "second sight." The capacity which he and his son Emile attained is so marvellous as to be worthy of record. For preliminary tests Houdin tried the estimation of the numbers of dots on dominoes till, he says, " we at length were enabled to give instantaneously the product [sum?] of a dozen dominoes. This result obtained, we applied ourselves to a far more difficult task, over which we spent a month. My son and I passed rapidly before a toy-shop, or any other displaying a variety of wares, and cast an attentive glance upon it. A few steps further on we drew paper and pencil from our pockets, and tried which could describe the greater number of objects seen in passing. I must own that my son reached a perfection far greater than mine, for he could often write down *forty* objects, while I could scarce reach thirty. Often feeling vexed at this defeat, I would return to the shop and verify his statement, but he rarely made a mistake."

This capacity may be compared with the results obtained by adults after practise for a month in the author's experiments: here the average number of objects named was but *six*, the maximum ten.

[2] A simpler, but less efficient, substitute may be contrived after the style of the 'krypteon' described by Sanford (6, p.403).

paper 10 × 15 cm., and that the exposure-card holder and the opening in the screen of the instrument are adjusted to corresponding dimensions. Collection of pictures of all kinds, cut from old magazines and trimmed to a size not over 10 × 15 cm. [Seconds' pendulum, Fig. 21, or metronome with electrical contracts (Fig. 14) and double time-marker, Fig. 20, in place of the 50-vibs. fork.]

Preliminaries.—Set up the tachistoscope as described in Test 24. Remove the fixation-card, or turn it off to one side. Replace the small by the large card-holder, and arrange the screen so that the rectangular replaces the square opening. Adjust the disc so that the pre-exposure section just covers the exposure-card when the instrument is set.[1] Adjust the weights and the sector of the disc to give a point-exposure of 3 sec. This will be secured approximately by setting the first weight at 10 cm., the second at 85.2 cm., and opening the sector 115 deg. For rough determination of the exposure, the stop-watch (Fig. 13) may be used. For accurate determination, use the seconds' pendulum or metronome and the signal-magnet (time-marker) mentioned in Test 10; make the determination as directed in Test 24.

The following *types of exposure cards* will be found desirable; 10–20 cards should be prepared of each type selected. (*a*) Groups of irregularly arranged *dots* or small circles or crosses, in number from 6 to 20.[2] These should be grouped in a space not exceeding 35 mm. square in the center of the blank, and care must be taken to avoid too obvious hints of combinations in the groupings. These cards may readily be made by the use of the asterisk (*) sign of a typewriter. (*b*) Cards on which *pictures*—drawings, cuts, lithographs, etc., cut from magazines—have been pasted. (*c*) Cards containing a single line of 8–10 three-letter *nonsense syllables*. These may be prepared most easily by the typewriter. As it is important to avoid syllables that resemble words when pronounced, a selected list is given here. The last line may be used to illustrate the effect of using sense syllables. (*d*) Cards containing varied combinations of lines in meaningless *drawings*. Make these with pen and ink; keep the drawings within an area 75 mm.

[1] Fixation is not so exact, of course, with this arrangement, but in view of the long exposures, the error may be neglected for comparative work.

[2] Cf. Aiken (1, p. 37) who terms this the test of "unconscious counting." Groups of about 14 dots prove satisfactory for testing adults.

high and 130 mm. long; use combinations of arrows, circles, loops, and straight lines, with 10–15 of these elements in each drawing. One or two drawings with some hint of meaning, such as a conventionalized desk-telephone, a 'woodeny' disjointed horse, etc., may be introduced if E desires to secure light upon S's use of verbal and other associations for holding the drawings in mind.

vel	ild	bli	ool	vid	elt	arl	ime	arg	tob
zen	rud	bri	euk	zet	ilt	omb	ull	ept	pha
vem	tud	rab	sef	dak	smi	irm	zin	ibe	urs
eit	urf	ift	ank	ung	nen	ruv	acq	ong	pru
nis	dri	aum	jur	geg	gla	euf	jek	pof	kun
zud	geb	yef	rik	tau	lud	gur	ige	rin	zib
dro	pum	rin	nuc	ahn	rad	ite	buh	orm	gos
wol	gah	orp	tef	uff	lel	pru	spo	eig	orl
zig	arb	pud	lom	teg	cha	baw	ipp	lin	ith
luh	zan	gom	fid	ruj	heb	ret	ume	ech	tas
tan	end	low	rat	not	eel	red	add	ton	arc

(*e*) Cards containing typewritten four-line *stanzas* from some not too difficult or too well-known poem. The object in using poetry is to secure a certain degree of equivalence in length, rhythm, style, and topic from one exposure-card to the next. (*f*) Cards containing columns of *digits*, as illustrated in Aiken, p. 30. A sample column from her book is the following:

230
729
11
36
40000
16
40

METHOD.—Proceed in general as in Test 24 (save that there is no preliminary fixation-mark). Inform S what *type* of card is to be exposed, *i.e.*, a drawing, a stanza, etc.

S's reports may be as follows: for *dots*, a statement of their number, supplemented, if desired, by a rough pencil sketch of their position and an account of their grouping; for *nonsense syllables*, *poetry*, or *digits*, an oral or written report; for *drawings*, a pencil

sketch; for *pictures*, a verbal description, supplemented, if S desires, by a pencil sketch. In the case of pictures, *E* may also quiz *S* with regard to the observation of various details which *S* has not reported, to see whether they have actually escaped his observation or have merely been neglected in his report, *e.g.*, by the use of such questions as: "Did you notice any details in the background?" "Is there any printing in the picture?" "What color was the girl's dress?" After *S*' report is finished, *E* may often obtain further light upon *S*'s work by confronting him with the stimulus-card and asking him what features or details he had failed to note, what he had forgotten, or what he had misapprehended.

Treatment of Results.—Some of the material lends itself well, some but poorly, to quantitative treatment, but the latter is often most useful for qualitative analysis of the mental processes concerned. The following system, though obviously arbitrary, has been found serviceable. (*a*) *Counting dots:* assign credit for the reporting of the correct number only; consider any mistake, even of one number, a failure. (*b*) *Pictures:* estimate *S*'s grade upon a scale of 10, *i.e.*, assign the grade 10 to a report which seems to indicate complete recall of all the salient features of the picture; assign the grade 0 to complete failure, and score intermediate grades accordingly. This grading may be made fairly objective by counting up the number of features or 'points' which are adjudged essential to a satisfactory report, and comparing the number given by *S* with this standard number for the picture in question. (*c*) *Nonsense syllables:* assign one unit to each letter correctly reported (*e.g.*, 4 syllables = 12). but deduct 0.5 for each error of transposition or insertion, whether of letters within syllables, or of the syllables themselves. (*d*) *Drawings:* as in the case of pictures, rate the reproduction on a scale of 10, by reference to the number of lines or elements correctly reproduced, in comparison with a standard number for the drawing in question. (*e*) *Poetry:* assign one unit for each word correctly reproduced, but deduct 0.5 for each error of transposition or insertion. (*f*) *Digits:* as with nonsense syllables, assign one unit for each digit correctly reported, but deduct 0.5 for each error of transposition or insertion, whether of digits within numbers or of the numbers themselves.

B. SIX SECONDS' EXPOSURE WITHOUT THE TACHISTOSCOPE

MATERIAL.—Small table. Piece of cloth, preferably gray, large enough to cover the table-top. Seconds' pendulum (Fig. 21). A piece of cardboard about 30 × 45 cm., and a full-sized sheet (22 × 28 in.) of gray cardboard. Collection of miscellaneous small objects, familiar enough to be named by all the *S*'s, *e. g.*, pencil, rule, spoon, tin box, leaf, cup, bunch of keys, toy animals, salt-shaker, postcard, etc. Ten different objects will be needed for each exposure.

METHOD.—Place the large sheet of gray cardboard on the table. Arrange on this as a background a group of ten objects, but avoid combinations of obviously related objects, such as pen and inkstand. Make a rough sketch of the group, with a list of the objects, so that it can be restored when desired for later tests. Cover with the gray cloth (to conceal the objects while *S* is taking his position and receiving instructions). Let *S* now stand in front of, and close to, the table, but in a position that will not interfere with its full illumination. Let him hold the smaller sheet of cardboard so as to cut off the view of the table-top. Start the seconds' pendulum, which must be placed somewhat to one side, where it can easily be seen by *E*, but will not distract *S*. Remove the cloth and take the cardboard which *S* has been holding. Give *S* a "ready" signal, and 2 sec. later, quickly remove the cardboard screen. At the expiration of 6 sec.[1] again cut off *S*'s view with the screen. Let him immediately turn his back to the table, and give a verbal description of the objects.

The chief stress should be placed on naming as many objects as possible: afterwards, *S* may be asked to describe the details of the objects or to make a rough sketch to indicate their relative positions. For qualitative purposes, *S* should be encouraged to give an account of the manner in which he observed the objects and the manner in which he has reproduced them. After *S* has enumerated as many objects as possible, exhibit the objects that were unnamed, either singly, or mingled with a number of objects not on the table, and ask *S* if he can identify any more of the objects exposed.

[1] The simplest method is to count the strokes mentally, "one, two, three," etc.: if the exposure is made at "one," the screen is to be restored, of course, at "seven."

TREATMENT OF RESULTS.—Credit *S* with one unit for each object named.

RESULTS FOR BOTH METHODS.—The following results are based on the author's trial of these tests with three adults, *G* and *R* (young men, experts in psychology), and *V* (a young lady school teacher, with a small amount of psychological training).

(1) The most striking quantitative result is the very small increase in the *range of apprehension in comparison with the range of attention* (Test 24). Thus, in an exposure of 10–50σ, an average *S* can grasp 4 or 5 objects: here, with an exposure more than 100 times as long, the average *S* enumerates but 6 objects (with a minimum of 3, and a maximum of 10). Similarly, 3 sec. exposure of nonsense syllables allows, on the average, 10.15 letters (Table 33), *i. e.*, between 3 and 4 syllables, to be read correctly, which is approximately the same as can be read with exposures of a small fraction of a second. On the other hand, the 3 sec. exposure of sense material gives an average "range" of nearly 12 words in contrast to the 4–6 word limit for ordinary tachistoscopy. This interesting advantage of sense material in the longer exposure is evidently due to the fact that such material can be grouped and recalled by larger and more meaningful units, whereas the heterogeneous combinations of nonsense syllables or disparate objects are more difficult to identify and recall. In the case of poetry, *S*'s feel that the limit of their performance is set simply by the amount that can be read during the exposure, whereas, even in the 6 sec. exposures, there is not time enough clearly to apprehend 10 disparate objects. The maximal reproduction of poetry was that of *R*, who read the first 19 words of the following, with 3 sec. exposure:

> "Were they unhappy then? It cannot be.
> Too many tears for lovers have been shed,
> Too many sighs give we to them in fee,
> Too much of pity after they are dead."

(2) *Individual differences* in capacity for quick apprehension are clearly indicated by Table 33. Thus, *V* excels in the estimation of dots and in reading poetry, but is the poorest *S* in reading nonsense syllables, in reproducing drawings, or in describing pictures and objects. *G* excels in these performances, but is handicapped in

TABLE 33

Individual Differences in Visual Apprehension (Whipple)

MATERIAL	DOTS*	PICTURES	NONSENSE	DRAWINGS	POETRY	OBJECTS
G..........	11/33	6.96	10.90	8.65	9.42	7.10
R..........	16/35	6.89	10.70	6.42	12.92	5.57
V..........	28/35	4.40	8.85	3.70	13.21	5.50
Average ...		6.09	10.15	6.26	11.83	6.03

* Efficiency is indicated here by the number of times the group of dots was correctly reported, *i.e.*, 11 out of 33 trials, etc. As *G.* missed two trials, the average cannot be figured exactly.

reading poetry by his relative unfamiliarity with English poetry (he is of German descent): his poor capacity in estimating dots cannot be explained.

These results indicate that it is not possible to assert that an *S* has a given grade of general ability of apprehension, or even of visual apprehension: rather, we must state that he excels in the attentive observation of pictures, of drawings, of words, or of certain kinds of objects, etc. This confirms in an interesting way the general verdict of experimental work that mental ability is narrow and specific: here, for instance, we find that *V* is more than twice as efficient as *G* in the quick perception of groups of dots, while *G* is more than twice as efficient as *V* in the quick perception of irregular drawings.

(3) The effect of *practise* is shown in Table 34. Here, each "period" represents the average of three exposures, usually one daily for three days. Inspection of these data gives little warrant for the belief that systematic practise would enable an adult *S*

TABLE 34

Effect of Practise upon Visual Apprehension. Average for Three Observers (Whipple)

PERIOD..........	1	2	3	4	5	6	7	8	9
Pictures	6.6	4.9	5.9	6.3	6.9	6.9	6.5	5.5	
Nonsense ...	9.3	10.6	8.4	10.8	11.7	10.6	10.4	8.9	9.2
Drawings	6.6	6.3	5.0	5.6	5.0	7.7	6.5	6.6	5.7
Poetry.......	10.7	11.5	11.3	10.8	13.0	12.5	13.0	11.7	
Objects	5.6	6.3	5.9	6.0	5.9	6.5	6.4		

markedly to improve his ability for quick visual perception. The tests with dots do not lend themselves readily to quantitative treatment. The seeming improvement with drawings during the 6th period was due to the use of one very easy drawing in that group. There is some slight evidence of an improvement in reading poetry, which amounts roughly to the addition of one word, but this may be attributed to increased familiarity with the peculiar style of the poem in use. If any improvement can be inferred in the case of the objects-test, it must amount, on the average, to the addition of less than one object.

What meager practise-effect is here to be detected is entirely explicable by increased familiarity with the material in use, and by the appearance of the little tricks or schemes of grouping mentioned below. It is highly unlikely that this period of practise really trained the attention of the *S*'s in such a way as to enable them to excel others in the quick observation of objects or incidents other than those used in the experiments. It is impossible to explain the marvelous exploits reported by Miss Aiken[1]—one sample of which may be cited here. The column of numbers that we have mentioned (p. 247) was exposed by Miss Aiken for 3 sec. only; "the pupils were then asked but once to multiply the first number by two, to extract the cube root of the second, to square the third, to extract the square root of the fourth, to divide the fifth by two, to multiply the sixth by twenty-four, and to divide the seventh by four, and then to repeat the changed column, which they did as follows: 460, 9, 121, 6, 20,000, 384, 10."[2]

(4) A *qualitative analysis* of the data secured in this experiment shows that the efficiency in visual apprehension exhibited by any *S* will depend on the following:

(*a*) *Native capacity* for concentrating attention in general. This is the factor which it is desired primarily to isolate and measure, but it is impossible to secure such a measurement until the other factors are measured or eliminated.

---

[1] From correspondence with some of Miss Aiken's pupils, I infer that it was her class as a *whole*, and not any single pupil, that was able to accomplish these feats. In other words, one pupil undertook one, another a second, another a third phase of the reproduction, etc. Even so, the performance is amazing.

[2] See Miss Aiken's book, pp. 30 ff. for other examples.

(*b*) The *degree of attention* given at the exposure in question. In theory, each exposure is accompanied by *S*'s maximal attention: in practise, this is not always secured. Tests in which *S* reports distraction must be thrown out. The effect of good preparedness on *S*'s part may be illustrated readily by exposing test-cards without warning of their type; the consequent elimination of 'expectant attention' will reduce *S*'s efficiency.

(*c*) *Individual capacity* of *S* to attend to, and to assimilate, the particular *type of material* in use—drawings, nonsense syllables, etc.

(*d*) The *ease of assimilation* of the particular test-card in use. Thus, an easy bit of poetry will increase the performance of all *S*'s; a drawing that can be *named*, however fancifully, can on this account be held longer and reproduced better by most *S*'s.

(*e*) *Obstruction or distraction:* Some feature in the object displayed, whether important or trivial, will often catch *S*'s attention, interfere with his exploitation of the balance of the exposure field, and thus measurably reduce his performance. Thus, a misprint in a line of poetry, or the presence of some unusual word, will induce most *S*'s to reread the line, even at the obvious expense of their record.

(*f*) *Ideational type:* Visually-minded *S*'s hold drawings, pictures, and objects by their visual appearance, and are inclined to use visualization for the reproduction of at least portions of the nonsense syllables and poetry. Auditory-minded *S*'s hold verbal material by auditory imagery: if decidedly auditory-minded, like the author, they may also attempt to hold even pictures, drawings, and objects in so far as possible in auditory terms by using verbal formulations, names, etc., as cues for recall.

(*g*) *Restriction:* *S*'s performance is definitely conditioned by his voluntary attempt to assimilate and reproduce either a large amount or a small amount of the material exposed. Thus, the nonsense syllables exposed are ten in number: by an effort, all ten may sometimes be read, but the result will be a poor reproduction of two or three syllables; if *S* confines his attention to the first four syllables, he may read these twice, and succeed fairly often in reporting all four. Similarly, there may be voluntary restriction of attention to other types of material.

(*h*) *Grouping:* As noted in Test 24, visual apprehension is greatly facilitated by any device that permits the grouping of the constituent elements in the exposure-field. This factor, more than any other, gives us the explanation both of the individual differences and of the practise-improvement above mentioned. Thus, in the dot tests, the mass of irregularly arranged dots is, by most *S*'s, arbitrarily rearranged (subjectively) into two, three, or sometimes four, groups of dots—each group containing three to six dots. Drawings are, similarly, often split up into component elements, and then recombined by a sort of analysis and synthesis. Sometimes this process is accompanied by the application of verbal symbols as tags for recall, *e.g.*, "a rectangle, two peaks, and an arrow." For most *S*'s, this analytic-grouping method turns out to be more effective than the 'steady stare' which they are prone to employ at first.

REFERENCES

(1) Catherine Aiken, Methods of mind-training, N. Y. 1896. Pp. 75. For supplementary exercises, see her Exercises in mind-training, N. Y.

(2) Life of Robert Houdin, the king of conjurers, by himself, Philadelphia, 1859. (Ch. 14, p. 256.)

(3) J. P. Hylan, The distribution of attention, in P. R., 10: 1903, 373–403 and 498–533.

(4) W. B. Pillsbury, Attention, London and N. Y., 1908. Pp. 330.

(5) J. Quantz, Problems in the psychology of reading, in P. R. M. S., 2: No. 1, 1897.

(6) E. C. Sanford, A course in experimental psychology, Boston, 1895 and 1898. Pp. 449.

(7) G. M. Whipple, The effect of practise upon the range of visual attention and of visual apprehension, in J. E. P., 1: 1910, 249-262.

TEST 26

Cancellation.—In order to measure degree of attention, several forms of test have been devised in which a continuous task is assigned under conditions such that maximal attention is demanded for the best work, and that any reduction of attention is reflected directly in the speed or accuracy of the work. Prominent among these tests is that which we shall term 'cancellation.' The essential principle is the crossing out of an assigned letter or letters, or word, or other symbol from a printed sheet. Investigators have modified the test in various minor details, have given it different names, and have used it for different purposes. Thus, its originator, Bourdon (4), used it with ten adults to measure "discrimination," Oehrn (7) proposed it, under the title "search for assigned letters," as a convenient test of attention for experimentation in individual psychology; Cattell and Farrand (5) introduced it into the series of tests of Columbia University students in the form of the "*A*-test" for "rate of perception," and Thorndike employed it later, together with what may be regarded as variations of it (the "*a-t* test," the "*e-r* test," and the "misspelled-word test") for various comparative studies, particularly for his examination of the mental traits of twins (12); Binet (3, 236 ff.) called it the "correction of proof" test, and used it to contrast the ability of intelligent and unintelligent children (as did Winteler, 14), to measure fatigue (as did Ritter, 8), and incidentally, to study their capacity to

break and form associations; Miss Sharp (10), from whom the term cancellation has been borrowed, tried the test in her studies in individual psychology with Cornell University students; Vogt used cancellation to examine attention and distraction; Judd in his *Manual* (6), has classed the test among those devoted to discrimination reactions; the author has tried the test as a measure of fatigue, and also has sought to determine its correlation with class standing.[1]

The results of the cancellation test are conditioned in the main by three methodological factors—(*a*) the number of letters to be cancelled, (*b*) the kind of text or material employed, and (*c*) the duration of the test.

(*a*) *Letters cancelled.* It is easily demonstrated that the ranks of a given number of *S*'s will vary somewhat according as a few (1–2) or several letters (4–6) are to be cancelled. Where but one letter has been used, this has commonly been the letter *a* (Bourdon, Cattell and Farrand, Wissler, Thorndike, Sharp) or occasionally *e*. In cancelling more than one letter, we find that Bourdon tried at different times, *a* and *i*; *a, r, i, s*; *a, e, l, t*; or *a, e, l, t, o, k*; Binet, *a, e, d, r, s*, also *i, o, l, f, t*, and *a, e, r, o, s, m*; Vogt, *l, n, s*; Winteler *n, s, t*, and *l, m, r, s, t*.

It is Binet's contention that, when but one letter is cancelled, *S*'s tend to work with approximately equal accuracy, but with varying speed, whereas, when four or five letters are cancelled, they tend to work with approximately equal speed, but with varying accuracy; he, accordingly, arranges the test in one or the other of these ways, as he wishes to measure either speed or accuracy. As will be shown later, this assumption is not strictly justifiable.

The cancellation of one letter is so easy that some investigators have tried other devices for complicating the task: Ritter, for example, had his *S*'s cancel every *r* with a vertical stroke and every article (in German texts) with a horizontal stroke, or, again, every *s* with a vertical and every preposition with a horizontal stroke: his idea that these two assignments would be equally easy and could thus be used interchangeably was not borne out in practise. Thorndike has used still another variation: mark every *word* (in a Spanish text) containing both *e* and *r*, or both *a* and *t*. The same investigator has used a three-minute test in which misspelled words in a page of narration are cancelled as rapidly as possible.

(*b*) *Material.* We may distinguish five kinds of material— (1) spaced prose, (2) unspaced prose, (3) disconnected words, (4) unregulated pied material, and (5) regulated pied material, with (*a*) unequal, or (*b*) equal, proportion of the constituent letters.

[1] In an extended study of mental and physical tests upon 50 grammar-school children, the results of which have not yet been published in full.

(1) *Spaced prose* is secured by taking any ordinary printed matter, but preference has been given to scientific (Binet) or philosophical (Sharp) texts, rather than to easily grasped material (Winteler).

(2) *Unspaced prose,* as employed by Miss Sharp, is secured by having a page of concrete description, or of more abstract material, printed without capitals, spacing, or punctuation. The following lines will serve to illustrate two such texts:

> theshoresoftheislandwerecompletelyfringedwithbushesand
> greatcarehadbeentakentopreservethemastheyansweredasascreen

> theproblemofphilosophyhasbeenineveryagetodeterminethe
> relationbetweenbeingandthoughtbetweensubjectandobjecteveryphil

(3) *Disconnected words* were used by Sharp only: the text is a series of sense words arranged in nonsense combination.

(4) *Unregulated pied material* may be formed, after the fashion of Bourdon and Vogt, by printing a page of Hungarian, Finnish, or other little known foreign language, and omitting all capitals, punctuations, and combinations that suggest English words. The same material can be secured more simply by letting the printer set up a page of small letters picked at random from a pile of mixed type.

(5) *Regulated pied material* has a distinct advantage over other forms of material, for in the latter it is impossible to know without laborious counting just how many of any given letter are found in the passage, and the same passage does not afford the same opportunity for the use of different letters. The texts which contain actual words, whether spaced or unspaced, also

THE A-TEST

OYKFIUDBHTAGDAACDIXAMRPAGQZTAACVAOWLYX
WABBTHJJANEEFAAMEAACBSVSKALLPHANRNPKAZF
YRQAQEAXJUDFOIMWZSAUCGVAOABMAYDYAAZJDAL
JACINEVBGAOFHARPVEJCTQZAPJLEIQWNAHRBUIAS
SNZMWAAAWHACAXHXQAXTDPUTYGSKGVKVLGKIM
FUOFAAKYFGTMBLYZIJAAVAUAACXDTVDACJSIUFMO
TXWAMQEAKHAOPXZWCAIRBRZNSOQAQLMDGUSGB
AKNAAPLPAAAHYOAEKLNVFARJAEHNPWIBAYAQRK
UPDSHAAQGGHTAMZAQGMTPNURKNXIJEOWYCREJD
UOLJCCAKSZAUAFERFAWAFZAWXBAAAVHAMBATAD
KVSTVNAPLILAOXYSJUOVYIVPAAPSDNLKRQAAOJLE
GAAQYEMPAZNTIBXGAIMRUSAWZAZWXAMXBDXAJZ
ECNABAHGDVSVFTCLAYKUKCWAFRWHTQYAFAAAOH

interfere with different S's to variable extents by the tendency to the appearance of apperceptive processes, akin to reading. The first user of the cancellation test, Bourdon, regretted that he had not made use of regulated pied material; so, too, Vogt. This regulation may be of two forms. In the first, a single letter is chosen as a test letter and this is introduced a given number of times in chance order with other letters which occur less frequently. This arrangement is best illustrated in the so-called "A-test," which is here reproduced. It contains 100 A's and 400 other letters, 16 times each, and is printed in 11-point capitals.

The A-test has three disadvantages: it is not devised to permit the use of any letters other than A; it does not lend itself easily to four-letter cancellation; in many instances, two or more A's are in juxtaposition, so that some S's may see and cancel several at once.

To avoid these difficulties, recourse may be had to the second form of regulation, viz: the use of each of the 26 letters the same number of times each. In constructing the forms which are prepared for this test, the printer has mixed together 100 types of each letter and set them in chance order, save where a given letter was repeated too closely.

(*c*) *Duration.* It is preferable, whenever possible, to have all S's cover the same amount of material and to secure the quantitative record in terms of elapsed time. But, for group work, it is necessary to set a given time for all S's and to secure the speed record in terms either of extent of material examined or number of letters crossed. Most E's who have experimented with small groups have chosen the first method: the actual time expended upon cancellation has ranged from 1 to 20 min., or even several hours. At Columbia, 1 min. is allowed for the A-test, 2 min. for the *a-t* test, 3 min. for the misspelled word test. When working by the individual method, the A-test takes, on the average, about 95 sec. Miss Sharp's tests demanded from 3 to 4 min.; Bourdon used 6 min.; Winteler 15 min.; Binet found that by extending the time to 10 or 20 min., he could get signs of fatigue. When such long times are used, E may follow the time with the stop-watch and, at expiration of each minute, check with a small horizontal underlining mark, the point at which S is then working. It was Oehrn's idea that the cancellation test might be extended through several hours in order to analyze such factors as praise, fatigue, ennui, spurt, warming-up, etc. This has been partially done by Vogt, who worked for an hour or more at a time, in 10 min. periods with 5-min. intermissions.

Four forms of the cancellation test are here described; the first two deal with the cancellation in a regulated pied text of one letter and of four letters, respectively; the third reproduces the word-cancelling tests, and the fourth the misspelled-word tests, as administered at Columbia University.

A. Cancellation of a single letter

Materials.—Printed form, containing 100 of each of the 26 letters arranged in chance order.[1] Stop-watch. Pencil.

Method.—Whenever convenient, work with one *S* at a time. Place the form before *S*, printed side down. Instruct him as follows: "When I say 'ready,' turn over this sheet of paper, begin at the first line, and mark every *a*[2] on the page like this." (Exhibit for an instant another sample form, already marked, to be sure that *S* understands the instructions.) "Mark as rapidly as you can, but try not to leave out any *a*'s." Give the command—"ready." Start the watch when *S* glances at the first line: stop it when he finishes the last line; record on the form the time, together with *S*'s name, the date, and other needed items.

If necessary to follow the group method, as in conducting class-room tests, it is a good plan to write a sample line on the black-board and show how the cancellation is to be done. Make sure that no *S* turns the paper before the signal, and that all start simultaneously. All *S*'s must cease work at the command "stop," and underscore horizontally the letter at which they were looking when this command was given. The exact time to be used must be determined by a few preliminary experiments; it should be such that the fastest *S* to be examined will not quite be able to finish: 2 min. is to be generally recommended, since very few adults can finish within that time.

B. Cancellation of four letters

Material.—As before, using the form beginning *zycu*.

Preliminaries.—With a typewriter, place at the top of each sheet the letters—*q r s t*.[3]

[1] This form begins *hplg*, etc.: if it is desired to repeat the test with the same letter on a new form, use the one beginning *zcyu*.

[2] The letter *a* is chosen merely because it has commonly been used. If it is desired to repeat the test with other letters, it would doubtless be preferable to select a number which have been shown to be of nearly equal legibility. For this, one may recommmed *m* and *w*, or the four letters, *q*, *p*, *b*, *d*. (See Sanford, 8.)

[3] Save for special purposes, the cancellation of four letters suffices as well as that of five or six to bring out the characteristics of the test. The letters *q*, *r*, *s*, *t* are selected because they form a combination fairly easily remembered, and embrace one letter projecting above, one projecting below the line, and two small letters.

METHOD.—As before, with the added explanation that the letters to be marked have been placed for reference at the top of the sheet. (*E* must, of course, also state the letters verbally beforehand.) The time will naturally be longer than before: 4 min. may safely be used in most group tests, as a competent adult takes nearly 5 min.

C. CANCELLATION OF WORDS—THE A-T AND THE E-R TEST

MATERIAL.—Printed page from a Spanish text.

METHOD.—Instruct *S* to mark with a horizontal stroke each *word* containing *both a* and *t*. Exhibit a sample page, or illustrate on the blackboard. Forewarn *S* that the words are in a foreign language. If the individual method can be followed, take the time for the whole sheet; if the group method is followed, allow 2 min.

For a second test, use fresh sheets of the same text, but substitute *e* and *r* for *a* and *t*.

D. CANCELLATION OF MISSPELLED WORDS

MATERIAL.—Two printed texts containing a large number of misspelled words.

METHOD.—Test each text separately. In each case, instruct *S* to mark with a horizontal stroke every word that is not spelled correctly. Take the time for the whole sheet, or allow 3 min.

TREATMENT OF DATA.—In these four tests, we meet the difficulty which often appears, of having two or more indexes of efficiency—in particular, one of speed or quantity of work done, and one of accuracy, precision, or quality of work done. For some purposes it is valuable to keep these indexes distinct, but more often it is desirable to unite them into a single index which shall express fairly, though perhaps somewhat arbitrarily, *S*'s net efficiency.

If, as advised, we work by the individual method, the problem is simplified, for the elapsed time yields a direct index of speed. If we work by the group method, we have two possible indexes of speed—the number of *letters marked,* or the number of *letters examined.* In general, the latter is the safer index, since we have no certain guarantee, at least in the word tests, that the various portions of the text are of equal opportunity or equal difficulty, and cannot, for example, assume with certainty that the marking of 30 letters or words indicates half-again as much ability as the marking of 20.

The conversion of the speed and accuracy indexes into a single index has been attempted in various ways. Binet, as we noted, suggests that accuracy may be ignored when one letter is cancelled, and speed when several are cancelled. Cattell and Fullerton (5) propose to convert accuracy to speed by adding to the obtained time, the time that would be required to mark the letters omitted or wrongly marked—this increment being determined by simple proportion, on the basis of the actual time and the number of errors, but this correction has often been omitted in the use of the *A*-test. (Neither Vogt nor Winteler considered accuracy in ranking their *S*'s, even though 3 and 5 letters were to be cancelled.) Miss Sharp, on the other hand, converts speed to accuracy on the assumption that "in a given individual maintaining a constant degree of attention while doing a piece of work, the percentage of error is inversely proportional to the time taken for the work." While these methods seem theoretically justifiable, the actual reduction is very large in the case of some *S*'s. Possibly an empirical corrective formula might be experimentally determined.

Ritter calls attention to the prime necessity of making clear to *S* at the outset whether it is speed or accuracy that is the more desired. This emphasizes once more the necessity of having a standard formula for the instruction of *S*'s and of using it without variation. Ritter's own system of correction was as follows: add 1 for the omission of 1 line, 2 for the omission of 2 lines, 3.5 for 3 lines, 5 for 4 lines, 7 for 5 lines, and 2 for each subsequent line.

The following treatment is advised for the letter-cancellation data. When possible, rank speed in terms of time; when this is not possible, rank speed in terms of letters examined (ground covered).

Compute the index of accuracy according to the formula

$$A = \frac{c - w}{c + o},$$

in which A = the index of accuracy,

o = the number of letters erroneously omitted,

c = the number of letters crossed,

w = the number of letters wrongly crossed.

To compute a single index of efficiency (net efficiency), use the formula

$$E = SA,$$

in which E = the desired efficiency index,

S = the speed index (ground covered),

A = the accuracy index,

or the formula

$$E = \frac{S}{A}$$

when S = the speed index (in terms of time).

To illustrate: a 12-year boy in the *qrst*-test covered 825 letters, crossed 40 (including 2 *w*'s), and omitted 78. By the formulas just given, we find S = 825, A = .322, E = 266. This instance is characterized by extreme inaccuracy, so that the degree of penalization seems perhaps over-great, but it represents as fair a method of computing net efficiency as can at present be devised.

It is often of interest to keep records of the number of omissions for each letter, as in this way, one may get comparative estimates of their difficulty.

For the word-cancellation, similar methods may be pursued: when the test is administered by the time-limit method, efficiency, for the sake of simplicity, may be taken in terms of the number of words marked, with a deduction of 2 for each wrongly marked word.[1]

The rank in the misspelled word test may, similarly, be best computed in terms of time, with a reduction for errors as described for the cancellation of letters.[2]

RESULTS AND CONCLUSIONS[3].—(1) *Dependence on age.* There is undoubtedly a general improvement with age both in speed and accuracy of cancellation, but we have no elaborate comparative study of different ages. Wissler's results for the *A*-test (15) show that Columbia seniors better their freshmen records: in 35 cases, there was an average reduction in the time from 105.4 to 88.9 sec., and in errors from 4.7 to 1.6.

(2) *Dependence on sex.* Wissler found that women surpass men in the *A*-test: 270 men averaged 100.2 sec. with 2.2 errors; 42 women averaged 91.2 sec. with 3.0 errors.

[1] This method is followed by Thorndike, who found such errors only in one of ten papers on the average.

[2] Thorndike neglected errors in this test, though they occurred in about one-third of the papers.

[3] Unless otherwise stated, these results are based on the letter cancellation tests. For illustrative records of the other tests, consult Thorndike (12, or 11, pp. 47-9.)

(3) *Dependence on letter cancelled.* According to Bourdon, letters whose form is simplest are oftenest omitted; with the letter *o*, 6 to 10 times as many errors were made by his *S*'s as with the letter *e*, but the author's comparative tests with the letters *c* and *o* do not substantiate these conclusions fully. Table 35 shows that in 40 consecutive tests, grouped by tens, *o* is distinctly easier to cancel than *c*, save in the first group; it may therefore be expected that it is easier to cancel than *e*, which, in practise, is often confused with *c*. In the *qrst*-test, 50 boys committed errors of omission as follows: *s*, 337; *t*, 561; *r*, 653; and *q*, 718. Here there is no relation whatsoever with legibility, since *s* belongs to the group of relatively illegible letters, and *q* to the group of very legible letters (Sanford, 9). We may conclude that when school children cancel these four letters in one test, it is the least often used letter that suffers most from omission.

TABLE 35

Effects of Letters and of Fatigue on Cancellation (Whipple)

LETTER	FIRST TEN	*m. v.*	SECOND TEN	*m. v.*	THIRD TEN	*m. v.*	FOURTH TEN	*m. v.*	AVERAGE
c.......	93.12	5.36	91.04	5.73	89.86	3.0	96.64	4.69	92.66
o.......	94.06	4.93	86.28	1.42	82.64	3.4	87.52	3.46	87.62
Average.....	93.59		88.21		86.25		92.08		

(4) *Dependence on the form of material.* The effect of the material used for cancellation is indicated by the results of Bourdon and of Sharp. Bourdon found that a change in size of type from one in which the small letters were 1.75 mm. high to one in which they were 1.25 mm. high made little difference; if anything, the smaller type proved better. The same investigator found that more letters are cancelled in nonsense than in sense material. Sharp found that the use of unspaced, in comparison with spaced, material reduced the speed, but also reduced the errors. The average time for seven *S*'s for a spaced text was 190 sec., for an unspaced 219 sec., and the average percentage of error was 9.6 and 4.58, respectively. The faster speed and poorer quality of work done with ordinary prose is plainly due to the tendency for the rise of meaning to act as a

distraction to the process of cancellation; the latter requires attention to individual letters, whereas the former (ordinary reading) proceeds naturally by the assimilation of words as wholes. As will be noted below, most *S*'s do not 'read' the text when cancelling letters in pied material.

(5) *Dependence on number of letters to be cancelled.* According to Bourdon, an increase in the number of letters to be cancelled causes a progressive decrease in the extent of material examined, but approximately the same number of letters is cancelled in a given time. This conclusion may be accepted as generally valid, but much depends upon the arrangement of the material. Bourdon is of the opinion that *S*'s that are accurate in cancelling one, are also accurate in cancelling several letters. The author's tests confirm this generalization in the main, as they show a value for this correlation of +0.38 (Table 37). With regard to speed, Bourdon merely states that some individuals slow in marking one letter prove fast in marking several; my results indicate, however, that the correlation between one- and four-letter efficiency is greater in the case of speed (+0.49) than in the case of accuracy (Table 37).

(6) *Relation of speed and accuracy.* Binet, as previously noted, assumes that speed tends to be equalized in marking four letters, accuracy in marking one letter. My tests with 50 boys (Table 36, last column) show that, while there is a tendency in this direction, it is not sufficiently pronounced to warrant the neglect of either speed or accuracy in estimating efficiency. In other words, the variation in speed is proportionately less in marking four letters than in marking one letter, and the variation in accuracy is proportionately less in marking one letter than in marking four letters, but there is nothing approaching equalization of either factor.

TABLE 36

Averages and Variations in Cancellation Tests (Whipple)

LETTERS CANCELLED	INDEX	AVERAGE	*m.v.*	COEFFICIENT OF VARIABILITY
e	Speed	702	193	27
q, r, s, t	Speed	811	177	22
e	Accuracy	89.6	8.0	9
q, r, s, t	Accuracy	63.1	11.4	18

Cattell and Farrand state that some S's are slow and accurate, some slow and inaccurate, some fast and accurate, some fast and inaccurate. Since, however, an S that works very fast presumably tends to work less accurately, we may expect to find indications of an inverse correlation between speed and accuracy, and this is the case. Wissler's relatively easy A-test gives an inverse correlation ($r = -0.28$). In the author's tests, the cancellation of one letter is harder than that of the A-test, and the cancellation of four letters is still harder; in consequence, we find inverse relations of -0.37 and -0.64, respectively (Table 37). In the case of 30 Cornell University students, speed and accuracy were found similarly inversely related, but by a lower coefficient ($r = -0.48$, Table 37).

TABLE 37

Correlations in Cancellation Tests: 50 Grammar-School Boys (Whipple)

FIRST MEMBER	SECOND MEMBER	PEARSON COEFFICIENT
Speed, one letter	Accuracy, one letter	−0.37
Speed, four letters	Accuracy, four letters	−0.64
Speed, four letters	Accuracy, four letters	−0.48*
Speed, one letter	Speed, four letters	0.49
Accuracy, one letter	Accuracy, four letters	0.38
Speed, one letter	Class-standing	−0.40
Accuracy, one letter	Class-standing	none
Net efficiency, one letter	Class-standing	−0.32
Speed, four letters	Class-standing	−0.40
Accuracy, four letters	Class-standing	0.39
Net efficiency, four letters	Class-standing	−0.09
Net efficiency, one letter	Word-building (Test 47)	none

* This correlation refers to 30 University students.

(7) *Dependence on practise.* Practise increases efficiency in all cancellation tests, as is illustrated in Tables 35, 39 and 40. Whether this practise-effect concerns only the letter used, or extends to all letters, cannot be stated until a series of equivalent letters has been determined by preliminary tests. Continued practise with the same letters almost doubles speed, and raises accuracy to a maximum. The letters are not held in mind by conscious effort but recognized quasi-automatically, and the whole process becomes unexpectedly simplified.

(8) *Dependence on fatigue.* The experiments of Bourdon, Binet, and Ritter show that cancellation is affected by fatigue, which reduces accuracy, rather than speed—a result in accordance with what we know of the effects of fatigue on other forms of mental activity. It follows that the cancellation of four or more letters is better adapted than the cancellation of one letter for testing fatigue. If but one letter is cancelled, practise and warming-up may easily conceal fatigue, as is shown in the author's continuous tests, extending over two hours and characterized by marked subjective fatigue (Table 35).[1]

In his examination of fatigue, Binet compared the first and second half of a 20 min. test, and found that his *S*'s had made 54 errors in the first, and 95 in the second, 10 min. Ritter was successful in 8 of 10 trials in getting indications of fatigue in school children by his form of the cancellation test: Table 38 gives a sample series of errors.

TABLE 38

Effects of Fatigue on Cancellation (Ritter)

TIME	PREVIOUS SCHOOL EXERCISE	ERRORS
9 a. m.		37
9.55 a. m.	Greek	94
10.10 a. m.	Pause for Vespers	78
12 m.	Livy and Chemistry	84

(9) The *correlation between cancellation and intelligence* is obscure. Wissler found no correlation between the *A*-test and class standing. Binet's results are gained from four different series: in Series 1 (cancellation of five letters), the intelligent and unintelligent showed little difference in speed, though the former increased their speed more in the second half of the test; the unintelligent, however, made four times as many errors. In Series 2, conducted 15 days later, the unintelligent equalled the intelligent in speed, but were still inferior in accuracy. In Series 3, a sudden change was

[1] It may be stated, however, that some three or four tests which followed the fourth group, but which are not included in the table, indicate a distinct increase in time required, *e.g.*, from 85-95 to 106-108 sec. The quality of the work showed no progressive changes.

made in the letters assigned; this reduced the speed of all *S*'s more than one-half, and the intelligent, rather oddly, made more errors than the unintelligent. Binet explains this on the ground that they, the intelligent, had established their associations more strongly in previous series.[1] In Series 4 (20 min., 6 letters, two of them new), the speed was about equal for the two groups, but the unintelligent made more errors.

Winteler's gross results (Table 39) indicate the superiority of his intelligent children, but when his *S*'s are ranked individually, two of the four unintelligent are found to be superior to some of the intelligent. Similarly, the unintelligent seemed as capable as the others in adapting themselves to the change from crossing three to crossing five letters, so that Winteler concludes that one cannot, on the whole, discern any inferiority on the part of the unintelligent with respect to the number of letters cancelled, to their quickness of adaptation, or to the steadiness with which attention is maintained within the series. It is to be regretted that Winteler did not take any account, direct or indirect, of the number of errors.

TABLE 39

Relation of Average Number of Letters Cancelled to Intelligence (*Winteler*)

LETTERS	*n s t*			*l m r s t*			
	FIRST DAY	SECOND DAY	BOTH	FIRST DAY	SECOND DAY	BOTH	ALL TESTS
Intell..........	277	329	606	312	414	726	1326
Unintell.......	255	303	558	248.5	326.5	575	1133

The author's own tests are summarized in Table 37: it is evident that there is an inverse relation (−0.40) between speed and class standing; that, when one letter is cancelled, there is no correlation, but when four letters are cancelled, there is a direct correlation (0.39) between accuracy and class standing: that, when accuracy

[1] It would be interesting to see whether, in such a test, the intelligent would surpass the unintelligent, provided the latter had, by added practise, been brought to an equal state of proficiency before the change of letters. In accordance with Binet's thesis that intelligence is indicated primarily by readiness to adapt oneself to a new situation, we should then expect the unintelligent to make the greater number of errors.

and speed are conjoined in a single index (net efficiency), there is a definite inverse correlation for one letter, and a possible inverse correlation for four letters, between such efficiency and class standing. In other words, the best pupils work more slowly at the cancellation test; if four letters are cancelled, this slower speed has its reward in a relatively high degree of accuracy.

(10) *Other correlations.* Wissler found no correlation between cancellation and reaction-time to sound, and only a low correlation (0.21) between the *A*-test and a test of quickness in naming colors. The same investigator found that weak eye-sight was conjoined with inaccuracy in the *A*-test, but that the reverse was not true.[1] Thorndike found a correlation between twins of 0.73 in the *A*-test, 0.75 in the *a-t* and *e-r* tests and in the misspelled word test.

The author's grammar-school boys showed no correlation between net efficiency in cancelling one letter and the word-building test.

Aikens, Thorndike, and Hubbell found no correlation between quickness in the misspelled word test and quickness in the *e-r* test, and a correlation of 0.16 (8th-grade pupils) to 0.25 (5th-grade pupils) between accuracy in these two tests, as measured by number of words or letters marked per line. The correlation of efficiency in the *e-r* test and in addition and association tests was also found to be slight or none, but efficiency in misspelled words and in addition and associations tested was correlated by 0.50.

(11) *Dependence on the form of movement.* Bourdon's description of the process of cancellation seems to imply that the examination of the line is interrupted during the process of cancelling itself. It is doubtless true that the movement of reaction does interfere sometimes with the process of perception and recognition. Vogt has attempted an experimental analysis of this interference by comparing the amount of ground covered in the usual method and the amount of ground covered when the assigned letters are simply recognized but not marked.[2]

[1] In certain series conducted by the author with a University student to test the relative values of various letters, the net result of two months' work was to indicate the probability that the student had astigmatism!—an inference which was confirmed by the oculist.

[2] There is here no check upon accuracy: it would seem better to let *S* name the letters as fast as recognized, or at least utter some simple sound or tap his finger at each recognition as suggested above.

He found that, in his own case, the marking slowed the speed of the performance as a whole by at least 15.8 per cent., in the case of another, less practised *S*, by the astonishing amount, 42.3 per cent. If these conclusions are generally valid, we shall have to admit that the apparently 'mechanical' phase of the cancellation work may easily be the determining factor in the individual differences which the test may reveal, and this will alter radically our conception of the nature of the test.

But, in the author's opinion, Vogt's results are so warped by the intrusion of practise and warming-up (*Anregung*) as to be inconclusive. They certainly do not accord with the results in Table 40, which represent the author's tests upon himself by several variant methods.

TABLE 40

Effects of Different Methods of 'Reaction' in qrst-Test (*Whipple*)

NUMBER	TIME ELAPSED SINCE PRECEDING TEST	METHOD OF REACTING	SPEED IN SEC.	OMITTED OUT OF 199*
1	————	Finger tapping	283	?
2	10 minutes	Electric key	246	3
3	10 minutes	Actual marking	207	3
4	15 hours	Electric key	220	0
5	5 minutes	Actual marking	193	4
6	5 minutes	Mere recognition	197	?
7	2 hours	Electric key	237	1
8	1 hour	Actual marking	184	4

* In the test in use, there were only 49 s's, owing to a printer's error.

In this work, the same text and the same letters were employed in eight trials on two days. Test 6 was in accordance with Vogt's series in which *S* merely recognizes the letters, and executes no movement of reaction. In Test 1, the finger was lifted slightly as each letter was recognized. In Tests 2, 4, and 7, this movement was changed into a simple tap upon an electric key, which was connected to an electric recorder: this device enables one to record the total number of movements of reaction and hence to measure accuracy in terms of omissions, while the movement is so familiar and simple as to be virtually negligible. In the remaining tests, the letters were cancelled by pencil strokes in the regulation manner.

It is evident that when two or three tests are administered in close succession, there tends to be improvement due to practise and warming-up, especially the latter. If the process of marking delayed the performance by any appreciable amount, we should expect Test 6 to be shorter than Test 5, and it would be difficult to account for the minimal time shown in Test 8. Moreover, the author can testify from introspection that there is no conscious delay introduced by the movement of cancelling itself. It is quite possible that some *S*'s may be delayed, however, by the cancelling, and it would be profitable, whenever time permits, to investigate, by appropriate tests, the nature and extent of this retardation in speed.

(12) *Qualitative analysis.* Most *S*'s do not pronounce the letters of the text, as is shown both by introspection and by the fact that a greater number of letters can be examined than could be read over silently in the same period, *e.g.*, 1876 and 1086, respectively (Bourdon). The letters to be cancelled, however, are often mentally pronounced by *S*, especially if four in number, in order to keep them in mind. Attention is then arrested by the sight of the assigned letters, which are recognized either visually or by inner pronunciation.

The most common error is that of omission. When four letters are marked, *S* often temporarily forgets one or more letters. Less often, *S* makes no marks at all for limited parts of a line, or even for whole lines. The latter defect is, in the author's experience, characteristic of either very young, or very careless *S*'s.

Adults may hit upon the device of traversing every other line from right to left; this seems to economize time and to insure at least as great accuracy.

REFERENCES

(1) H. Aikens, E. Thorndike, and E. Hubbell, Correlations among perceptive and associative processes, in P. R., 9: 1902, 374–382.

(2) A. Binet, Attention et adaptation, in A. P., 6: 1900, 248–404.

(3) A. Binet, L'étude expérimentale de l'intelligence, Paris, 1903. Pp. 309.

(4) B. Bourdon, Observations comparatives sur la reconnaisance, la discrimination et l'association, in Rev. philos., 40: 1895, 153–185.

(5) J. Cattell and L. Farrand, Physical and mental measurements of the students of Columbia University, in P. R., 3: 1896, 618–648.

(6) C. Judd, Laboratory manual of psychology, N. Y., 1907. Pp. 124.

(7) A. Oehrn, Experimentelle Studien zur Individualpsychologie; in P. A., 1: 1896, 92–151.

(8) C. Ritter, Ermüdungsmessungen, in Z. P., 24: 1900, 401–444.

(9) E. C. Sanford, The relative legibility of small letters, in A. J. P., 1: 1888, 402–435.

(10) Stella Sharp, Individual psychology: a study in psychological method, in A. J. P., 10: 1899, 329–391.

(11) E. L. Thorndike, An introduction to the theory of mental and social measurements, N. Y., 1904. Pp. 210.

(12) E. L. Thorndike, Measurements of twins, in Archives of Phil., Psych, etc., No. 1, Sept., 1905 (Columbia Contrib. to Phil. and Psych., 13: No. 3). Pp. 64.

(13) R. Vogt, Ueber Ablenkbarkeit und Gewohnungsfähigkeit, in P. A., 3: 1901, 62–201.

(14) J. Winteler, Experimentelle Beiträge zu einer Begabungslehre, E. P., 2: 1906, 1–48, 147–247.

(15) C. Wissler, The correlation of mental and physical tests, P. R. M. S., 3: No. 6, 1901. Pp. 62.

TEST 27

Counting dots.—This test was devised by Binet (1) and employed by him and later by Winteler (2) in their comparative studies of intelligent and unintelligent children in order to measure the degree of attention. The problem set before *S* is that of counting a number of dots which are arranged in an irregular group or in lines of varying length and spacing.[1] When this work is attempted without the aid of pointer or pencil to keep the place, it is distinctly difficult and necessitates active concentration, but by selecting different arrangements of dots, this difficulty can be graded to suit *S*'s of different degrees of development, and comparative scales of ability or norms for given arrangements of dots may presumably be established for each age.

Material.—Stop-watch. Two sets (in duplicate) of 27 printed test-cards.

These cards are numbered in the upper left-hand corner in accordance with the following plan: Cards A 1 to A 10 contain rows of dots with uniform spac-

---

[1] The test evidently has similarity with the dot-counting under 3 sec. exposure (Test 25), but the removal of the time restrictions and the increase in the number of dots make the conditions quite different: here it is the degree rather than the range of attention that is primarily to be measured.

ing (for each card): Cards Ba 1 to Bd 4 contain lines with groups of 2, 3, 4 or 5 dots each in which the spacing within the groups and between the groups varies as indicated in Table 41: Cards C 1 to C 5 contain 5 arrangements of dots in irregular clusters. These three kinds of material reproduce those found to be of value by Binet and Winteler. The dots, like Winteler's, are 1.5 mm. in diam.[1] The 'A' cards were used by both experimenters, the 'B' cards by Winteler only, the 'C' cards by Binet only. The term 'points' in the Table refers to the printer's point or typographic unit: one point is 1/72 inch.

TABLE 41

Specifications for Test-Cards Used in Dot-Counting

CARD NUMBER	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
Points in Spacing	20	19	12	10	8	5	4	3	3	3
Number of Dots	13	15	14	17	21	27	29	45	52	60

CARD NUMBER	Ba1	Ba2	Ba3	Ba4	Bb1	Bb2	Bb3	Bb4
Points within Groups	6	6	6	6	4	4	4	4
Points between Groups	18	18	18	18	12	12	12	12
Number of Dots	45	47	55	52	50	54	70	64

CARD NUMBER	Bc1	Bc2	Bc3	Bc4	Bd1	Bd2	Bd3	Bd4
Points within Groups	3	3	3	3	3	3	3	3
Points between Groups	9	9	9	9	15	15	15	15
Number of Dots	70	73	81	90	65	75	70	72

Card Number and Number of Dots: C1, 50; C2, 48; C3, 49; C 4, 54; C5, 57.

METHOD.—Make the tests in the order indicated by the card-numbers. Instruct S: "Find the total number of dots on this card: count aloud, in any way you wish (that is, by ones, twos, threes, etc): work as rapidly as you can, but try particularly to get the number right." The emphasis is thus placed on accuracy rather than upon speed. E should record S's time, but without the latter's knowledge. He should also keep before him a duplicate of the card upon which S is working and should note thereon S's method of grouping, in order to discover whether he counts always by ones, or always by twos, etc., and whether he accommodates

[1] Binet found that S's with poor eye-sight had difficulty with the test when the dots were 1 mm. in diam.

himself to the objective grouping of the 'B' cards. (*S*'s attention should not, of course, be called by *E* to the grouping in these cards.) *E* should also record the magnitude of the error and its nature—whether overestimation or underestimation—but should not communicate this information to *S*.

Treatment of Results.—Binet and Winteler both ranked *S*'s merely in terms of accuracy and put no time-limit upon their performance. It would seem possible, after some experience, to discover the relation between speed and accuracy, and possibly to make use of a corrective formula as in the case of other tests where these two factors appear, *e.g.*, the Cancellation Test (No. 26). The errors are to be counted simply by subtracting the given from the true number, *e.g.*, 62 for 65 represents 3 errors.

Results.—(1) Many *S*'s have a *constant error*, but this may be either an error of overestimation, or an error of underestimation.[1]

(2) Winteler found that some *S*'s always counted by the same number, *i.e.*, used the same *increment in adding*, as "one, two, three, four," etc., or "two, four, six, eight," etc. When the change was made to the 'B' material, he found that all three of his bright children and one of the dull adapted their counting to the objective grouping, whereas the rest of the dull children continued for the most part to employ the form of counting (almost invariably by ones or by twos) that they had adopted in the 'A' cards.

(3) In examining the *relation between dot-counting and intelligence*, Binet concludes that, although the test exacts a high degree of attention, the outcome depends more upon *S*'s care than upon his intelligence. His results were confused by the presence of one bright child with poor eye-sight: when this case is eliminated, the intelligent children are found to make fewer errors than the unintelligent: in one series the relation is 13 to 19, in another 16 to 24; the difference, as in many other tests, tends to lessen with practise.

Winteler likewise found the bright children, as a group, more

[1] No attempt has been made to analyze the conditions under which these constant errors appear; they might conceivably be due in part to illusions of filled and empty space, in part to individual differences in method of keeping the place in the line, in part to temperamental attitudes (overcautiousness, careless haste, etc.). This test, like many others, has not been subjected to careful introspective analysis by trained adults. But critical qualitative study of this sort is as desirable for the intelligent employment of any test as is the mere accumulation of quantitative results.

accurate, and also more rapid, than the dull children, but there were individual exceptions. The average for 10 series on each of two days were: errors, bright 8.33, dull 17.25; time in sec. per 10 dots, bright 4.97, dull 6.30—a suggestion that the tendency to inverse relation of speed and accuracy may not be so clearly evident as in many other tests. The difference in capacity of the two groups becomes striking, however, when the 'B' type of material is employed, because the unintelligent, as already noted, fail to adapt themselves to the objective grouping and make a large number of errors (79.75 to 9.00 of the intelligent).

(4) The outcome of the test is not affected by general ability in arithmetic; some of Winteler's *S*'s who did poorly had good grades in arithmetic.

REFERENCES

(1) A. Binet, Attention et adaptation, in A. P., 6: 1899 (1900), 248–404.

(2) J. Winteler, Experimentelle Beiträge zu einer Begabungslehre, in E. P., 2: 1906, 1–48, 147–247.

TEST 28

Reading complicated prose.—Reading, like counting, is a form of perceptual activity that has been proposed as a means for measuring attention. In reading, as in counting (Test 27), the process has been complicated in some manner in order that the increased difficulty may exact a higher degree of attention and so furnish a better opportunity for the study of individual differences. Miss Sharp, for instance, followed this plan when she sought to test degree of attention by requiring *S*'s to read two texts (a page of concrete description and a page of abstract exposition), which were printed without capitals, punctuation, or spacing.

In the present test, this plan has been extended, first by printing the complicated text backward, as well as without spacing, and second, by adding a test-sheet of similar subject-matter and identical length, but of regular form, in order to supply a check-test of maximal speed of reading under normal conditions.

MATERIALS.—Two printed texts: (*a*) a page of prose in regular form, (*b*) a page of equivalent, but 'complicated' prose. Stop-watch.

METHOD.—The regular text (*a*) is used first. *E* gives the following directions. "When I say 'now,' I want you to read this aloud, just as fast as you can without making mistakes." *E* records the time; also, if desired, the number of errors. He does not, however, correct *S*'s errors in this part of the test.[1]

In using the reversed and unspaced text (*b*), *E*'s directions are: "When I say 'now,' read this aloud as fast as you can without making mistakes. You will find this page more difficult than the one you have just read, because you will have to begin here in the lower right-hand corner and read it backward, and because this page is printed without any punctuation or capitals and without spaces between the words. I shall not give you any help, but if you make a mistake, I shall stop you, and ask you to correct it."

E follows *S*'s reading upon a duplicate text. He records the time for the entire reading, and by glancing at the watch at every pause in the reading, he notes upon the duplicate text at the points concerned, the time in sec. consumed by *S* at these pauses. These notations should be made for every pause of 5 sec. or over. In case *S* pauses for 30 sec. at any point, *E* then supplies for him the word or phrase which he needs to continue his work.

To secure accuracy, *E* must correct every error in *S*'s reading, even slight errors, such as singular for plural forms, etc., and he must especially avoid the temptation to assist *S* whenever he halts, save for the 30 sec. halts, as just stated. He must notify *S* of each error, as it occurs, by simply interjecting 'no,' and must indicate its place by repeating the two or three words just preceding it. For example, if *S* reads: "they were all alike in tone," *E* interrupts with: "No!—all alike in?—" *S* corrects himself: "all alike in one respect," etc.

VARIATIONS IN METHOD.—In order to measure, and to be able to allow for, individual differences in maximal rapidity of articulation, *E* may require *S* to reread the normal text four or five times,

[1] With most adult, the errors are few in number and trivial in nature. To attempt to correct them would render it impossible to measure the speed of reading. If the errors are numerous, *E* may, however, ask *S* to reread the text, this time without making a single error, but still as fast as possible. The slight advantage that *S* gains by knowing the subject-matter is of little moment compared to the false advantage that he has gained by hurrying his reading so fast as to commit many errors.

or until the subject-matter is thoroughly familiar and further repetition fails to reduce the time. A brief rest-pause should follow each trial to avoid cumulative fatigue. Another method is to test *S*'s time for counting aloud to 50 at maximal speed, though here it is often more difficult to check or control *S*'s tendency to gain time by slurring or otherwise mutilating the words he pronounces.

RESULTS.—(1) The results obtained from a limited application of the reading test are summarized in Table 42, in which three groups of *S*'s are represented—(*a*) five dull children aged 12.5 to 16.75 years, (*b*) five bright children aged 10.33 to 12.75 years, and (*c*) 26 students in university and summer school classes—including two university instructors.

TABLE 42

Results of Reading Tests, in Sec. (Whipple)

GROUP	NUMBER	READING FORWARD				READING BACKWARD			
		Aver.	M. V.	Min.	Max.	Aver.	M. V.	Min.	Max.
Dull Children	5	116.4	19.	101	169	1061.	291	814	1500*
Bright Children	5	100.0	11.6	85	125	544.4	167.2	490	910
University Students	26	73.3	8.5	47	100	320.	100.	125	755

* This time is estimated from the amount of the text covered in 10 minutes.

(2) A comparison of the performance of the bright pupils with mature students shows that *age* exercises an appreciable reduction in the time needed to read prose aloud, particularly when the task is artificially complicated.

(3) A comparison of the performance of *the dull and the bright* children of approximately the same school grades shows the clear superiority of the bright pupils, despite the fact that they are some two years younger than the dull pupils. Here, again, the difference is accentuated in the complicated text.

(4) Individual differences are more pronounced when the reversed text is used.

The coefficient of variability, for example, in the group of adults is approximately 11 per cent for forward, as contrasted with 33 per cent for 'backward' reading.

The actual range of performance is also surprisingly large. Thus, in the adult group, the quickest backward reading is only 1/6 as long as the slowest backward reading. Miss Sharp's seven S's showed even greater individual differences—ranging from 143 to 900 sec. for concrete, and from 125 to 405 sec. for abstract texts. When it is remembered that her S's were all college students in advanced classes, the variability in performance seems unexpectedly large, and it is hard to understand Miss Sharp's declaration: "We had expected to discover individual differences of much more definite character and much greater amount."

(5) The speed of reading the complicated text correlates in a rather high degree with the speed of reading the normal text: for the adult group cited in Table 42, $r = 0.79$, $P.\ E.$ about .09.

This relation appears, to the author at least, unexpected, and hence of special interest. Adults who try the test are almost unanimous in their declaration that their speed in reading forward is not conditioned by the task of assimilating the substance of the text, but solely by the physiological limit to intelligible articulation: their speed in reading the reversed text, however, is quite obviously not conditioned by speed of articulation, but by a sort of 'linguistic readiness,' or ease of apperceiving the constituent words or phrases of the text. If these statements are correct, we are evidently driven to the conclusion that persons who read difficult and complicated subject-matter rapidly also tend to speak more rapidly—a conclusion that subsidiary tests appear to confirm. It is, furthermore, not unlikely that fast readers are also fast thinkers as well as fast speakers, though this generalization has at present no experimental verification.

NOTES.—We have not had sufficient experience as yet with the complicated, or reversed prose test to understand fully the nature of the processes upon which it depends. The considerations just developed make it evident that these processes embrace something besides attention, if, indeed, attention plays any large share in the conditioning factors. For fast readers, in the reversed text, the proper combinations 'rise up' like the hidden faces in the puzzle picture when once they have been seen.

Since a facile apperception of printed symbols would appear, on theoretical grounds, to be a natural concomitant of good intelligence, it is possible that this test may prove to have considerable

diagnostic value. To determine this point, however, it needs extended trial with control both by introspective analysis and by the statistical examination of all possible functional correlations.

REFERENCE

Stella E. Sharp, Individual psychology: a study in psychological method, in A. J. P., 10: 1899, 329–391.

TEST 29

Simultaneous adding.—In Tests 24 and 25, the attempt was made to measure the field of consciousness or range of attention during a relatively brief period. Tests have been proposed in which attention is solicited by several claimants, not for a brief period, but continuously. These tests may be grouped as tests of simultaneous activity. Their primary purpose is to ascertain how successfully a number of activities can be carried on at the same time. Ordinarily, disparate activities are selected (Test 30). Less often, as in the present case, a single type of activity is complicated or 'spread' in such a manner as to demand simultaneous attention to more than one phase of activity.

In simultaneous adding, as devised and conducted by Binet in his comparative study of six bright and six dull children, the task is to carry on a series of additions in three columns of figures at once.

MATERIALS.—Prepared forms, ruled in series of three vertical columns, with the numbers 6, 28, 43, printed at the head of the first three columns. Stop-watch. Pencil. A piece of cardboard about 20 cm. square.

METHOD.—Instruct *S* that he is to continue for six minutes, as fast as he can, adding *one* to each number, and writing the sum directly below. Thus, he first writes: 7, 29, 44, then 8, 30, 45, etc. Exhibit a sample form on which such additions have been made. The moment that *S* has written a line (three sums), cover it with the cardboard, so that he must hold the three sums in his mind from one line to the next.

Record the number of additions made in 6 min., and note the number and nature of the errors.

VARIATIONS OF METHOD.—For mature *S*'s, the addition of 1 to each sum may not be difficult enough: *E* may then complicate the

task by requiring *S* to add, say, 3 to the first, 1 to the second, and 2 to the third column. If desired, repeat the test with another variation in the constants to be added, say, 2 to the first, 3 to the second, and 1 to the third column.

Results.—(1) Binet found that this test excited a great deal of interest, and exacted a greater effort of attention than any he tried.[1]

(2) There are marked differences in the *amount of work* (number of additions) done by different *S*'s, *e.g.*, in Binet's tests, from 40 to 96 numbers, *i.e.*, from 13 to 32 lines of 3 numbers each.

(3) Examination of the *errors* shows that *S*'s commonly center their attention upon either one or two columns: here they make few or no errors, but the neglected columns contain many errors. In other words, errors in all three columns at once are rare. Binet's pupils wrote 100 of 245 lines erroneously; of these 100 errors, 59 were found in one column only, 34 in two columns, and 7 in all three columns. Errors are more frequent in the third column than the second, and in the second than the first; by count, 76, 60, and 31, respectively.

(4) Contrary to the results of many laboratory tests, says Binet, the number of errors committed tends to increase with continued work. This, we may surmise, is due in part to the fact that the later portion of the work necessarily deals with higher numbers, and in part to the confusion and loss of confidence that is felt after a number of errors have been made.

(5) *Degree of intelligence* appears to have little influence upon this test. Binet states that speed is not at all a matter of intelligence, merely an individual variation. The unintelligent make more errors than the intelligent, but the difference (17 *vs.* 13.4) is not as great as one might expect.

Notes.—As a matter of fact, it is improbable that simultaneous adding really exacts a 'spread' of attention to all three columns. What *S* does is to write a given line on the basis of his memory of the preceding line: his additions are made successively; he is not really adding three columns at once. The test, therefore, really measures what is termed "immediate memory" rather than the

[1] For detailed illustrative work secured from his *S*'s, consult his article, p. 384.

spread of attention. The results may also be conditioned by the readiness with which *S* handles numbers and performs additions generally.

REFERENCE

A. Binet, Attention et adaptation, in A. P., 6: 1899 (1900), 248-404.

TEST 30

Simultaneous disparate activities.—In tachistoscopy and visual apprehension (Tests 24 and 25), we measure the range of attention for simultaneous *impressions*: in simultaneous adding (Test 29), we test the capacity of attention for concurrent activities of a *homogeneous* type: in the present test we employ *disparate activities*, and study what Meumann terms "heterosensory" distribution of attention (7, i., p. 503). Theoretically, the measure of the capacity of an individual to direct his attention to the execution of several activities at once should be of importance, since this capacity seems to imply the possession of such traits as keen concentration, mental alertness, quick-wittedness, and general intelligence. Gifted men, like Napoleon and Caesar, are said to have possessed this capacity in high degree. The latter, for instance, could dictate four letters while writing a fifth.

There seems to be a possibility that such a distribution of attention may take place under some circumstances, at least a distribution to two lines of activity, but strict experimental examination of the phenomenon is not easy, because, in practise, one of two things usually occurs; if both activities are difficult, attention alternates between them, and the activity not attended to at any moment is temporarily reduced, if not altogether suspended: and if one activity is relatively easy, it becomes, after short practise, reduced to automatism, so that attention can be given freely to the other. It is, then, difficult, if not impossible,[1] so to arrange experimental conditions as to secure *continuous division* of attention to disparate activities. On the other hand, it may be said that the very capacity to alternate attention quickly and successfully from one activity to another, or to reduce one activity quickly to autom-

[1] Cf. the remark of Titchener (11, p. 376): "Simultaneity of two psychologically disparate 'attentions' is, in my experience, altogether impossible."

atism, is itself an indication of important capacities—particularly of well-trained, highly concentrated attention, quick adaptability, and general mental alertness.

The test selected, simultaneous reading and writing, is but one of a large number of possible combinations, others of which are suggested below. This test has been proposed by Paulhan (8), by Binet and Henri (2, pp. 446–7), and tried in several forms by Miss Sharp (9) as a test of "range of attention."

MATERIALS.—A selected poem—preferably one which is divided into a number of stanzas of equal length—which will be of interest to *S*, but which is not well-known to him. Stop-watch. Pencil and paper.

METHOD.—(1) Let *S* read aloud at his *normal rate* a given section (about 8 lines) of the poem.

(2) Let *S* read another section[1] of the same length, and while reading, write the letter *a* as many times as possible. Continue the test by the use of other sections of the poem combined with the writing (3) of *a b*, (4) of *a b c*, and (5) of the entire alphabet.

This last test is the most satisfactory and should be the one employed if time permits but a single trial. It is important that *S* should *try* to maintain his reading at the normal rate.

VARIATIONS OF METHOD.—(1) Let *S* read both passages at his maximal instead of at his normal rate.

(2) Repeat the test several times with fresh texts to determine the effect of practise upon its performance.

(3) Compare the effect of striving especially for a large number of letters written with the effect of striving especially for a normal or for a maximal rate of reading.

TREATMENT OF RESULTS.—In practise, it will rarely be found that *S* maintains his normal rate of reading, particularly when writing the whole alphabet. To avoid the difficulty of working with two quantities, rate of reading and number of letters written, it is desirable to reduce these to a single "index of simultaneity." This is done, as is illustrated for the whole alphabet test in Table 43, by subtracting the normal reading-time from the reading-time during simultaneous activity, and dividing the number of letters

[1] Sharp used prose, and had her *S*'s read the same section five times. This has the disadvantage of tending to automatic reading. The advantage of securing identical length is practically assured here by the uses of stanzas of poetry.

written by this difference.[1] Thus, observer *B* read normally in 28 sec., with alphabet-writing in 113 sec.—a difference of 85 sec. He wrote 91 letters, and has an index of 1.07.

RESULTS.—The results for Miss Sharp's seven *S*'s are summarized in Table 43. It will be noted, (1) that the reading-time is usually lengthened by the complication of writing, (2) that more letters can be written with three than with two letters and more with two letters than with one letter, but not fully three times and two times as many, (3) that the writing of the whole alphabet is much different; either the reading is very much slowed, or fewer letters are written than when only *a* is employed, (4) that the *S*'s differ markedly in their capacity to carry on two processes simultaneously. The rank of the *S*'s in this test did not, however, correlate with their rank in any other of Miss Sharp's tests.

TABLE 43

Simultaneous Reading and Writing. (*Sharp*)

S's	TIMES OF THE FIVE READINGS IN SECONDS					NUMBER OF LETTERS WRITTEN				REDUCTION IN 5TH READING	INDEX OF SIMULTANEITY 5TH TEST
	1st	2d	3d	4th	5th	*a*	*ab*	*abc*	Alphabet		
B.....	28	38	42	50	113	47	62	78	91	85	1.07
G.....	22	22	22	21	28	29	34	39	40	6	6.66
V.M..	29	30	30	30	50	40	56	57	46	21	2.19
W.M..	26	27	27	27	29	27	28	36	13	3	4.33
E.R..	27	27	29	27	31	31	40	48	20	4	5.00
L.R...	22	25	26	25	37	41	44	51	26	15	1.76
T.....	27	29	30	31	29	36	40	45	25	2	12.50
Average..	26	28	29	30	45	36	43	51	37	19	4.78

(5) Paulhan noted that the simultaneous performance of two relatively easy activities did not take as long as the performance of the two in succession. He says: "I write the first four verses of *Athalie,* whilst reciting eleven of Musset. The whole performance occupies 40 sec. But reciting alone takes 22 and writing alone 31,

[1] Miss Sharp divided the difference by the number of letters. The reverse procedure has the advantage of indicating the degree of simultaneity directly, as a large quotient means good ability.

or 53 altogether, so that there is a difference in favor of the simultaneous operations.[1]" And again: "I multiply 421,312,212 by 2; the operation takes 6 sec.; the recitation of 4 verses also takes 6 sec. But the two operations done at once only take 6 sec., so that there is no loss of time from combining them."

NOTES.—Several *other tests* of a similar nature may be briefly described; still others may be contrived by *E* to suit conditions.

(1) As suggested by Meumann (p. 504), the *Cancellation Test* (No. 26) may be combined with other forms of activity, *e.g.*, let *S* cancel one or more letters and at the same time repeat short sentences read to him by *E*, or listen to the reading of a page of narration (Cf. Test 39) and repeat as much as possible of it after the cancellation is finished, or discriminate two points on the skin (Test 23), etc.

Vogt (12) combined the cancellation of three letters in a nonsense text with reaction to *metronome-beats* in the following manner: the metronome was set at 38, and the bell attachment set for every other stroke, so that there were 19 bell-strokes per minute; in some series *S* was required to make a slight movement of the finger at every bell-stroke, in other series also to lift two fingers at every fourth bell-stroke. Vogt found that this 'metronome-counting' retarded the total process of cancellation from 11.6 per cent. to 35.2 per cent., but that it did not affect appreciably the simple apprehension of the letters without actual cancellation (see Test 26, Result 11); in other words, he concludes that the movements of reaction to the metronome interfered with the movements of reaction in cancelling, but did not interfere with the apprehension of the letters in cancelling. This result is difficult to interpret if we do not admit Vogt's contention that the marking is in itself an appreciable factor in the cancellation test.

(2) McDougall (6) has proposed a form of 'dot tapping' to test the capacity for continuous exertion of attention in the following manner: Place upon a kymograph drum a sheet of white paper on which have been printed eight rows of 120 red dots; each dot is 1.5 mm. in diameter, and 5 mm. distant vertically from the next in the row; each series of 120 dots is arranged in an irregular line,

[1] The 'telescoping' may amount to more than this in the case of some individuals: see Titchener, 11, p. 375.

which covers an extreme width of 10 mm.. but the displacement of adjacent dots is not more than 5 mm. in the horizontal direction. This zigzag line of dots is now viewed, as the drum revolves, through a horizontal slit 10–15 mm. in the vertical dimension, and somewhat wider than the row of dots. *S* tries to strike each dot with a blunt soft pencil, and the drum is rotated at a speed (about one rev. in 23 sec.) such that he can succeed in striking each dot only by maximal effort. *S*'s work is graded as follows: for the omission of a dot or the making of an extra mark, count 1 error; for each lateral deviation of more than 1 mm., or each vertical deviation of more than 2 mm., count 1/2 error. Sample records show 50 to 150 errors in a series of eight rows, *i.e.*, 960 dots. For simultaneous activity tests, require *S* to undertake some other work at the same time, *e.g.*, mental arithmetic, reaction-time with the left hand, esthesiometry, etc.

In the few trials that the author has given this test, there has appeared a decided tendency for the dot-marking to lapse into automatism.[1]

(3) Both Binet (1) and Jastrow (5) have tested the *interference of intellectual processes with simple motor activities*.[2] To repeat these experiments, close one end of a relatively soft-walled rubber tube; connect the other to a Marey tambour (Fig. 17) and adjust the tambour for a graphic record upon the kymograph. Let *S* press or pinch the tube either (*a*) at an optimal rate, (*b*) at a maximal rate, (*c*) in groups of 2, 3, 4, or more pressures with stated time-intervals between the groups, *e.g.*, 3 quick pressures per sec., (*d*) in alternate groups of fours and sixes, etc., (*e*) in time to the beat of a metronome, or (*f*) in time to a melody which he himself hums, or in

[1] For a modification and development of McDougall's method with particular reference to the determination of degrees of clearness in attention, see Geissler (4, pp. 515 ff.).

[2] Burnett (3) has recently suggested a test in which visual attention is measured under conditions of visual distraction. Two mazes are employed, which are alike in every respect save one. Each maze is an ink line drawn in an irregular, wandering way over a white paper surface about 18 x 26 cm. In the second maze, small, embossed pictures and bits of paper of various forms and colors are scattered thickly among the twistings of the maze, though not actually covering any part of it. In use, the maze is covered with a glass plate, *S* is instructed to trace the pattern of the maze accurately and as rapidly as possible with a small wooden pointer. The measure of attention is afforded by the comparison of the time taken in Maze 1 (without distraction) with that taken in Maze 2 (with distraction).

any manner that will provide a suitably complex task'[1] Meanwhile, let him read sentences or disconnected words either silently or aloud, or let him undertake the mental addition of two-place numbers.[2]

The *general results* of such tests are: (1) the amount of interference of the two activities is proportional to their complexity and general difficulty; (2) movements that involve counting are more disturbed by adding than by reading; (3) reading or adding aloud interferes more with motor activity than does reading or adding silently; (4) the reading of disconnected words is more easily interfered with than the reading of sentences; (5) additions are slower and less accurate when performed with, than when performed without, motor activity of the 'tapping' variety; (6) concurrent intellectual processes affect the motor activities mentioned by (*a*) lengthening the interval between pressures, (*b*) diminishing their recorded height, (*c*) confusing their number or arrangement, or (*d*) causing the appearance of various motor incoördinations, tremblings, unevennesses, etc., which may amount well-nigh to a 'motor delirium;' (7) *S* may or may not be conscious of these disturbances in his motor activity; in general, he can give but obscure or fleeting attention to the pressures if the mental task is at all difficult; (8) the experiment soon induces symptoms of fatigue; (9) individual *S*'s differ noticeably in the degree of complexity of the motor action that they can execute successfully while engaged in intellectual activity—differences which appear to depend primarily upon the extent to which the motor activity may be reduced to automatism.

(4) Binet suggests a number of methods for testing ability to execute *concurrent motor activities*, which may, with a little ingenuity, be turned to account in the arrangement of simple tests; *e.g.* (1) make with the right hand a circular movement parallel to the median plane of the body in a clock-wise direction and with the left

[1] One might, for instance, adopt the plan suggested by Squire (10) for measuring fatigue of attention. Use the tube and tambour; let *S* memorize a series of eight or ten digits, *e.g.*, 6,9,2,1,3,6,4,7, and then tap this 'pattern' as rapidly as possible. Introduce concurrent processes and study their effect upon the tapping.

[2] An excellent method is to give *S* two numbers to start with, and instruct him thereafter to add at each addition the larger digit in the previous sum, *e.g.*, if 16 and 8 are assigned, the correct series will be—16,24,28,36,42,46, etc.

hand a simultaneous movement in a parallel plane in the reverse direction; (2) duplicate the registering apparatus above described, so as to provide a tube for each hand, and require *S* to press regularly and rapidly with the right hand, but to press with the left hand only twice for each five pressures of the right hand: (3) take a fountain-pen or pencil in each hand; with the right hand write some familiar poem and simultaneously with the left hand describe a series of small circles, or make a series of *u*'s with the right, and a series of dashes with the left hand. In the last named test there will be seen, as a rule, a tendency toward the production of similar movements, *i.e.*, the dashes become *u*-like, or the *u*'s spread out in a dash-like fashion. If *S*'s attention be called to this tendency, he may inhibit it by active control, but the tendency will usually recur the moment his attention becomes distracted—an observation that suggests the possibility of securing in this manner an index or measure of active attention.

REFERENCES

(1) A. Binet, La concurrence des états psychologiques, in Revue philos., 29: 1890, 138–155.

(2) A. Binet and V. Henri, La psychologie individuelle, in A. P., 2: 1895 (1896), 411–465.

(3) C. T. Burnett, A new test for attention against distraction, in P. B., 7: February, 1910, 64.

(4) L. R. Geissler, The measurement of attention, in [illegible]P., 20: 1909, 473–529.

(5) J. Jastrow, The interference of mental processes,—a preliminary survey, in A. J. P., 4: 1891, 219–223.

(6) W. McDougall, On a new method for the study of concurrent mental operations and of mental fatigue, in B. J. P., 1: 1904–5, 435–445.

(7) E. Meumann, Vorlesungen zur Einführung in die experimentelle Pädagogik u. ihre psychol. Grundlagen, Leipzig, 1907. (2 vols.)

(8) F. Paulhan, La simultanéité des actes psychiques, in La Revue Scient., 39: 1887, 684–9.

(9) Stella Sharp, Individual psychology: a study in psychological method, in A. J. P., 10: 1899, 329–391.

(10) Carrie R. Squire, Fatigue: suggestions for a new method of investigation, in P. R., 10: 1903, 248–267.

(11) E. B. Titchener, Lectures on the elementary psychology of feeling and attention, N. Y., 1908. Pp. 391.

(12) R. Vogt, Ueber Ablenkbarkeit u. Gewöhnungsfähigkeit, in P. A., 3: 1901, 62–201.

CHAPTER VIII

Tests of Description and Report

The two tests which are described in this chapter have certain features in common which demarcate them, on the one hand from the tests of perception and attention of the previous chapter, and on the other hand from the memory tests of the succeeding chapter, though, in many other respects, they resemble these tests.

The essential idea in both of the present tests is to determine capacity, not merely to attend and observe, or to recall what has been observed, but to put the results of this observation into linguistic form. If the observer gives his account of the experience at the time of his observation, this constitutes description; if at some time subsequent to his observation, this constitutes report.

It is evident that this giving of an account of an experience, particularly if the experience be somewhat complicated in form, is a more complex psychical process than those under discussion in the tests of attention and perception. This greater complexity makes the reduction of the observer's performance to exact quantitative terms a matter of greater difficulty, but, on the other hand, the activity called forth is more akin to that demanded in everyday life, and it is for this reason that these tests have been felt to possess a peculiar value, particularly in the study of individual differences in mental constitution and mental efficiency. Again, language occupies so strikingly prominent a place in our mental economy that tests which seek to bring out the observer's ability to cast experience into linguistic form are, on that account, well worth while. This is particularly the case in the second form of test, that of the report, which, in connection with the "psychology of testimony," has of late had a prominent place in psychological research.

TEST 31

Description of an object.—The description test first came into prominence through the work of Binet, who urged that the study

of individual psychology may be best advanced by resort to the experimental examination of complex, rather than of simple mental processes, and who considered the description test of special value in this connection. Binet made preliminary tests with Henri in 1893 (3), and worked at the test later by himself (1, 2). His method has been followed, though not in exact detail, by LeClere (5), Sharp (7), and Monroe (6).

MATERIALS[1].—Ordinary cigarette of the "Sweet Caporal" type. Cancelled 2-cent postage stamp.[2] Lithograph, entitled "Hindoos."[3]

METHOD.—(1) For the picture-test, supply *S* with writing materials; place the lithograph upright before him, about 75 cm. distant. Instruct him: "Write a description of this picture so that one who had never seen it would know all about it." Allow 10 min.

(2) For the cigarette-test, give the following instructions, and no others: "I'm going to put on this table before you a small object. I shall leave it there under your eyes. I want you to write a description of it; not to draw it, but describe it in words. You will have about 5 min. Here is the object." If *S* is busy at the end of the allotted time, or has written but a few lines, the time may be slightly extended.

(3) For the stamp-test, proceed in a similar manner, save that *S*'s are not forbidden to draw the stamp, if they wish to. The in-

[1] The cancelled stamp was used by Monroe, the cigarette by Binet. The lithograph is substituted for the different pictures that have been used by other investigators (Binet and Henri used Neuville's "The Last Cartridge," Binet a picture representing Fontaine's "Le Laboreur et ses Enfants," Miss Sharp "The Golden Wedding" and "The Interrupted Duel"), because of the impossibility of securing these particular pictures, or of the difficulty of using them under the conditions that prevailed in the original experiments (Binet's school children were well acquainted with the fable from Fontaine, for example).

If it is desired to extend the list of materials, *E* may employ other objects used by Binet (2), such as a box of matches, a penny, a leaf, etc.

For group tests, there should be at least one picture for every 5 *S*'s, one cigarette for every 2 *S*'s, and a stamp for each *S*.

[2] For group tests, it would be desirable to secure a set of stamps whose cancellation marks were approximately the same. The stamps should be trimmed off in such a manner as to show the full border of the stamp and a narrow margin of the paper upon which it was attached.

[3] This picture is one of a series called Leutemann's Types of Nations, catalogued by E. Steiger & Co., New York. It may be purchased, like all other materials cited in this book, of C. H. Stoelting Co., Chicago, Ill.

structions may run: "Describe this postage stamp so that a person who had never seen one would know all about it." Allow 10 min., or more if needed.

Treatment of Data.—In general, the results of the description test are not intended to be submitted to exact quantitative treatment, but are to be inspected for the purpose of forming an opinion of *S*'s general mental type and capacity. The papers may, however, be treated quantitatively, by (1) counting the number of words written, or (2) counting the number of lines written. *E* may, further (3), record in general terms the readiness and ease with which *S* undertakes the description, and (4) may rate his paper as a whole, with respect to its comparative merit, on a score of 10 for a satisfactory or adequate description. (5) The description may, perhaps, be classified also with respect to its general type or character, following the classification adopted by Binet, Le Clere, and others as explained below. (6) Descriptions of the postage stamp may also be catalogued with respect to the items mentioned, as was done by Monroe.

Results.—(1) The description of an object is inadequate, because it is almost invariably *simplified*, *i.e.*, a considerable number of its features, even important features, are unmentioned. Thus, in one of Binet's photographs, of the 22 objects or features that were mentioned at all, only 9.4 were mentioned, on the average, in each description.

(2) This simplification or reduction in the description is the result of what might be termed a process of *selection*. Certain features are mentioned in practically all descriptions, others are mentioned only occasionally. By tabulating the number of times each feature is mentioned, one may discover some of the principles which condition this selective process. Thus, in Binet's picture of the "Laborer," the old man is mentioned 36 times, his sons 30, his bed 29, the seated woman 27, etc., until we come to relatively unimportant objects that may almost escape mention at all, *e.g.*, a stick in the hands of one of the children—only 4 times in 36 descriptions. When pictures are used, persons are more often mentioned than furniture or other details of the setting of the scene.

Similarly, in the stamp test, tabulation indicates, according to

Monroe, the following order of frequency of mention: (1) word-inscriptions, (2) color, (3) number-inscriptions, (4) portrait, (5) substance, (6) form, (7) use, (8) perforated edge, (9) size, (10) cancellation, (11) ornamentations. The item *use* declines with age: all others are mentioned more frequently as age increases.

(3) *Sex differences.* Monroe states that girls generally mention more items than boys, and "*seem* to surpass boys in their knowledge of the postage stamp." It is not clear, however, whether this seeming superiority is due to better observation, to greater industry, or to greater zeal and conscientiousness.

(4) *Individual differences.* In 150 accounts of the photograph, Binet found no two alike. This wealth of individuality makes the description-test at once valuable and difficult—valuable as an indication of the variety of mental constitution, difficult as to quantitative or comparative treatment. As an extreme illustration, one may contrast the following descriptions of a postage-stamp—the first by a girl of 8, the second by a boy of 16.

(*a*) "The postage stamp has a picture in it. The postage stamp costs two cents. It says united states postage on it. The man has hair braided in back of his head. The Boarder is round. It has arms on it. The shape is square. The color is red. The man is White. You can get these to the postice [post-office] for two cents. There are lines around the boarder. The back of the stamp is white. It has nomber 2 on each side of it. The man has long hair."

(*b*) "COMMENTS ON THE ACCOMPANYING U. S. OF AMERICA 2 CENT POSTAGE STAMP.

"1. Its meaning: The Postage stamps have glorious history. In the past 57 years they have been more and more useful until now they are not only absolutely necessary, but constitute one of the great helps in the study of Geography, and one of the noblest pleasures for thousands and millions of people; Kings and Queens as well as children in the most miserable social condition.

"2. This Postage Stamp has the red color and is now next to the one penny stamps of Great Britain the most extensively used stamp used in the world. If I am not wrong its circulation in the past and present is the next largest of all others. The one penny stamp, I think has the first place.

"3. Its surroundings are very interesting. It is mounted on a piece of paper, remainder of an envelope, which fact easily indicates that it is used in the most cases for letter correspondence. I notice...... [Continues in this and the next paragraph a description of the stamp itself.]

"5. Some particular observations. I had 500–600 of them at home which my cousin had the kindness to send me. Of course they are of no special value, but yet they teach my little brothers the important lesson that such a little thing, like a stamp, will do all the necessary things for the transportation of a letter or other mail matter from the Atlantic to the Pacific. It is very interesting to me that with the march of civilization the great Postal system of the World has increased its actions more and more until it is now one of the chief functions under the sun. How much this single stamp has done I cannot say, but I know that some stamps, precisely like this, have done great service to the country."

(5) *Types.* Notwithstanding this diversity, investigators have sought to classify descriptions into a limited number of types. Thus, Binet proposes four types—the descriptive, the observational, the emotional (poetic, imaginative), and the erudite—each present in varying shades and degrees.

(*a*) The *describer*, or enumerator, as one might term him, merely catalogs the features of the object before him, with little regard for their interrelations, or for the meaning of the object as a whole.

Example: "The cigarette has the general form of a cylinder, cut at one end by an inclined plane where the paper is folded. It is stuffed with a rather dark brown tobacco. The paper is striped lengthwise. The paper is somewhat bruised. The tobacco projects about 0.5 centimeter from one end."

(*b*) The *observer*, though not necessarily more intelligent or clever than the describer, places more emphasis upon the interrelations of the several features that he mentions, interprets what he sees, conjectures and indicates the significance of the object as a whole. This type is also mentioned by Mrs. Bryant in her 'description-of-a-room' test (4).

Example: "A long, white, round object, composed of a paper cylinder, about 1/2 or 3/4 centimeter in diameter, filled with what is probably Oriental tobacco. It is about 7 centimeters long and must weigh about 6 grams [really 2 g.]. It is a badly rolled, uneven cigarette, and has been handled since it was pasted. In two places, to the right and left of the middle, the paper shows streaks as if it had been twisted. Other horizontal depressions indicate that there has been some pressure exerted upon the cigarette. I don't see the line where it has been stuck, but it must be badly fastened."

(*c*) The *emotional*, imaginative, or poetic *S* is less accurate in observation, but introduces emotion, sentiment or imaginative interpretation in his description.

Example: "It is a cigarette. It is thin, long, somewhat wrinkled. Its shape suggests a kind of elegant ease. Is it the cigarette itself, or the memories that it awakes that remind me somehow of a scape-grace? The cigar-

ette, there, all by itself on the table, makes me think of the bad student that goes off in the corner by himself to smoke. But I must write about the cigarette itself, and banish the idea of the smoker," etc.

(*d*) The *erudite* *S* tells what he knows, what he has been taught, or interjects bits of personal information about the object. This may indicate the presence of an unusual fund of information, or it may indicate sheer laziness, in that it is often easier to write what one knows than actually to describe from direct inspection.

Example: "We have before us here a cigarette. Let us see how it is made, In the first place, the exterior envelope is of light paper, called silk-paper. Then, inside is the tobacco. Tobacco is a product that grows almost every. where in warm or temperate climates. The leaves of this shrub are gathered- and, after a treatment which lasts four years, are turned over to the public in the form of powder, that is, snuff, or in shreds, as in the present instance," etc.

Miss Sharp did not attempt a classification into types, but noted that *S*'s observation "may be primarily directed to the particular objects or details of the picture, to the general arrangement of the objects, that is, the composition of the picture, or to the meaning of the picture, the story which it conveys,—the details observed being such as lead up to this interpretation, or explain and apply the interpretation that is given first. The different ways in which the same picture appeals to the various individuals indicate differences in mental constitution."

The results of LeClere's test are not directly comparable with those of other investigators, because his instructions were not to describe the object (gold watch), but to "write something that comes into mind as you look at it." He distinguishes in the contributions made by 30 girls, aged 13 to 17 years, seven types, viz: description, observation, imagination, moral reflection, erudition, pure or simple emotion, and esthetic emotion. He does not find, however, that any one of his *S*'s contributes a paper that may be classified in any one of these types, nor does any paper give evidence of a 'complete mind,' in the sense that all seven of the types are represented therein. In general, older or relatively more intelligent children write more varied or complex papers, *i.e.*, approach the theoretically 'complete' type of mental constitution.

NOTES.—The attempt to use the description test for classification of *S*'s into types of mental constitution is of obvious interest: the drawing of inferences from such a classification as to the mental make-up of the *S*'s is as obviously hazardous, for *S* may write his description in the vein that he thinks is wanted by *E*. Thus, Binet had reason to think that several *S*'s that he had classed as poetic or emotional were actually, in their everyday life, of a very matter-

of-fact and unsympathetic disposition. In general, the drawing of inferences from the work of *S's* would become safer in proportion as the descriptions were increased in number and variety, *i.e.*, an *S* who wrote in an emotional vein in four descriptions of four different objects has, presumably, a real emotional constitution.

REFERENCES

(1) A. Binet, Psychologie individuelle. La description d'un objet, in A. P., 3: [1896] (1897), 296–332.

(2) A. Binet, L'étude expérimentale de l'intelligence, Paris, 1903. Pp. 309.

(3) A. Binet and V. Henri, La psychologie individuelle, in A. P., 2: 1895, (1896), 411–465.

(4) Sophie Bryant, Experiments in testing the character of school children, in J. of the Anthrop. Inst. of Great Britain and Ireland, 15: 1886, 338–349.

(5) A. LeClere, Description d'un objet, A. P., 4: 1897 (1898), 379–389.

(6) W. Monroe, Perception of children, in Pd. S., 11: 1904, 498–507.

(7) Stella Sharp, Individual psychology: a study in psychological method, in A. J. P., 10: 1899, 329–391.

TEST 32

Fidelity of report (*Aussage* test).—Capacity to observe, or range of observation, may be tested by methods previously described (Tests 25 and 31); native retentiveness or capacity for recall may be tested by methods such as those that are described in subsequent sections; capacity to describe what is seen may be tested as has been indicated in Test 31, but there exists a type of activity, that of reporting a previous experience, which in a way combines these several activities, in that it demands both attentive observation, retention, recall, and an ability to marshall and formulate the items of experience in a verbal report (*Aussage*). In studying the 'psychology of testimony,' interest has been developed of late in the direct examination by experimental methods of the capacity to report, itself, and it has been found that reports may exhibit varying degrees of fidelity or reliability, more or less independently of the capacity that the reporters possess to observe or to retain experience; in other words, discrepancies or inadequacies may appear in reports, which are due, not only to misdirected attention, malobservation and errors of memory, but also to lack of caution or of

zeal for accurate statement, to scanty vocabulary, to injudicious phraseology, or, of course, to deliberate intent to mislead.[1]

Historically, the idea of subjecting capacity of report to test seems first to have been definitely proposed by Binet (1). Since then, the study of the psychology of testimony has found its most enthusiastic and active expositor in Stern, who has written an extensive monograph (15) on the subject, and in whose periodicals (*Beiträge zur Psychologie der Aussage* and *Zeits. f. angewandte Psychologie*) most of the work of subsequent investigators has, directly or indirectly, appeared. The applicability of this line of work to many practical problems, particularly in the field of jurisprudence, is too obvious to need further comment.

GENERAL METHODOLOGY OF THE REPORT-TEST

1. *Choice of material.* Of the several types of material that have been elaborated for the study of the report, *e.g.*, the picture-test, the event-test, the rumor-test, etc., the first mentioned has many advantages for our present purposes.[2] Two types of picture-test are prescribed; the first closely patterned after that employed by Binet in his study of suggestibility in school children, the second more in accord with the stock picture-test, as developed by Stern, Borst, Wreschner, Lobsien, and others.

2. *Choice of exposure-time.* For pictures, times ranging from 5 sec. to 7 min. have been used, though 45–60 sec. is most usual.

[1] It is true that no hard and fast line can be drawn between the report-test and the test of range of apprehension, or between it and the ordinary memory-test; in the main, however, range of apprehension implies a brief exposure followed by simple enumeration of the objects seen, so that what is tested is capacity to grasp or observe, rather than capacity to retain or to formulate. And the stock memory-test measures the *amount* of material that can be reproduced; here the learning is usually by heart, and the reproduction is largely mechanical. In the report-test, the object is more complex, the time of scrutiny much longer than in the observation-test, while stress is placed as much upon quality as upon quantity of reproduction, especially upon the fidelity of reproduction as conditioned by such personal factors as timidity, cautiousness, assurance, skill in verbal formulation, etc. Nevertheless, in the interrogatory, the report-test does closely resemble an ordinary test of memory.

[2] For a discussion of these advantages, of the several methods in detail, of the chief results, and for a general review of the whole field of the psychology of testimony, the reader is referred to an earlier discussion by the author (17). Suggestions for further tests will likewise be found therein.

The principle which has controlled the choice of exposure-time for the two tests that follow is to select such a period as will permit an average *S* to examine each detail of the object once.

3. *Choice of time-interval.* For the sake of brevity, the instructions that follow prescribe a report directly after the exposure. If circumstances permit, *E* will find it of interest to extend the interval to several minutes, or even hours or weeks. The effect of lengthening time-interval has not as yet been satisfactorily determined.

4. *Choice of form of report.* There are two distinct forms of report, (1) the 'narrative' (*Bericht, recit*), (2) the 'interrogatory' (*Verhör* of Stern, *Prüfung* of Wreschner, *interrogatoire* of Borst, *forage de memoire* or *questionnaire* of Binet).[1] The narrative is a free account, delivered by *S*, either orally or in writing, without comment, question, or suggestion by *E*: the interrogatory is a series of prearranged questions; the replies to these questions constitute the 'deposition' (*Verhörsprodukt*). The constituent parts of the narrative or the deposition may be termed 'statements' or 'items.' Each form of report has its advantages and its disadvantages: both should be employed whenever possible.

5. *Choice of form of interrogatory.* An interrogatory is 'complete' when its questions cover all features of the experience exhaustively, and are propounded to all *S*'s in the same order and manner: an interrogatory is 'incomplete' when its questions are restricted to such as refer only to those items not mentioned by *S* in his narrative. The interrogatories that follow are designed to be complete, but *E* may, by appropriate selection, convert them into the incomplete type.

6. *Choice of questions.* The form of questioning very materially affects *S*'s deposition, particularly if the questions are of the type known as 'leading' or 'suggestive' questions. If we follow Stern, at least six types of questions may be framed, viz: determinative, completely disjunctive, incompletely disjunctive, expectative, and consecutive.

[1] The terminology of the report-experiment has developed in Germany and France. I have been obliged to coin English equivalents—a task not always easy because the foreign terms have not been chosen with special care to secure consistency or to accord with legal phraseology. For this reason, the foreign equivalents are included here, and elsewhere in the discussion.

A *completely disjunctive* question is one that forces the reporter to choose between two specified alternatives, *e.g.*, "Is there a dog in the picture?"

An *incompletely disjunctive* question is one that offers the reporter a choice between two alternatives, but does not entirely preclude a third possibility, *e. g.* "Is the dog white or black?" In practise, for many reporters, especially for children, this form is virtually completely disjunctive, since a certain amount of independence is demanded for the choice of the third possibility, *e.g.*, for the answer "The dog is brown."

An *expectative* question is one that arouses a moderately strong suggestion of the answer, *e. g.*, "Was there not a dog in the picture?" (This is the form used by Binet to induce moderate suggestion.)

An *implicative* question is one that assumes or at least implies the presence of a feature that was not really present in the experience, *e. g.*, "What color is the cat?" In practise, it is clear that a determinative question might become implicative if the reporter had completely forgotten the item to which it referred. (The implicative question was used by Binet to induce strong suggestion.)

The *consecutive* question is any form of question that is used to augment a suggestion that has been developed by previous questions.

7. *Choice of method of grading. Treatment of data.* In general, the adequacy of a report depends both upon its quantity and its quality: quantity is measured by the number of items mentioned or the number of questions answered (in absolute or in relative terms) and is referred to as the range of report (*Umfang, étendue*): quality is measured by the fidelity of the statements made, and is referred to as the accuracy of report (*Treue, fidélité*).

We have also at our command useful indications of the positiveness or degree of assurance that S places in his report. Besides (1) complete uncertainty ("I don't know" or "I have forgotten"), we may distinguish (2) hesitancy ("I think" or "I believe"), (3) positive statement or assurance of ordinary degree, and (4) attestation or attestable assurance, *i. e.*, the highest degree of assurance, as indicated by S's willingness to take his oath that the statement is correct.

On this basis, the data may be subjected to treatment for the computation of a number of 'coefficients of report,' by the aid of the following simple formulas:

Coefficients of Report.[1]

Let P = number of possible items,
n = number of items reported (or replies made),
c = number of items reported with certainty (including attestation),
a = number of items whose correctness is attested under oath,
$n(N)$ = number of items reported in the narrative,
$n(D)$ = number of items reported in the deposition,
$n(r)$ = number of items that are rightly reported,
$c(r)$ = number of items that are certain and right,
$a(r)$ = number of items that are attested and right,
$a(w)$ = number of items that are attested and wrong,

Then

(1) n = range of report, absolute (*Umfang, étendue*),
(2) n/P = range of report, relative,
(3) $n(N)/n(D)$ = spontaneity of report,
(4) $n(r)/n$ = range of knowledge (*Umfang des Wissens, étendue du savoir*),
(5) $n(r)/c$ = accuracy of report (*Treue, fidélité*),
(6) c/n = assurance (*subjective Sicherheit, assurance*),
(7) $c(r)/c$ = reliability of assurance (*Zuverlässigkeit der Sicherheit, Sicherheitsberechtigung, fidélité de la certitude*),
(8) $c(r)/n$ = warranted assurance (*Sicherheit der Person, assùrance justifiée*),
(9) $c(r)/n(r)$ = accuracy of assurance (*justesse certifiée*),
(10) a/n = tendency to oath or attestable assurance (*tendance au serment*),
(11) $a(r)/n$ = warranted tendency to oath (*tendance au serment véridique*),

[1] The fourth formula is used by many writers, in place of the fifth, for accuracy of report; as here indicated, however, the indeterminate cases ("I don't know") are omitted from the denominator in computing accuracy. Next to range and accuracy, the most important coefficient is probably warranted assurance (8th formula), as a high ratio indicates a good witness, who reports a large number of items both correctly and with assurance.

(12) $a(w)/n$ = unwarranted tendency to oath (*tendance au faux-témoignage*),
(13) $a(r)/a$ = reliability of oath (*fidélité du serment*),
(14) $a(w)/a$ = unreliability of oath (*infidélité du serment*).

The determination of P, and hence of relative range of report, is often beset with difficulty: the most practical working rule is to rank as 'one item' any combination of features that forms a single natural working group, the details of which would escape individual observation under ordinary conditions. Or, again, P may be taken as the number of separate items mentioned by a competent S in describing the picture or test-object by direct observation.

Although different errors unquestionably have different degrees of importance (to forget a man is more serious than to forget the color of his necktie), no satisfactory plan for arbitrarily 'weighting' different items has been devised.

The psychologically best method of grading is unquestionably to classify the data statistically according to various categories—such as persons, objects, colors, sizes, etc.—and to compute range, accuracy, assurance and the other coefficients for each category separately. This will greatly increase the labor of quantitative treatment, but it will afford valuable insight into the qualitative conditions of report that could not otherwise be secured: the several coefficients can, for comparative purposes, be united subsequently into a single series of coefficients for the person or persons under consideration.

A. REPORT-TEST WITH BINET'S CARD OF OBJECTS

MATERIAL.—Rectangular sheet of orange-yellow cardboard, 33.5×40.5 cm., to which are attached two photographs, a label, a button, a penny, and a postage stamp.[1] Watch.

METHOD.—Give S the following instructions: "I want to try an experiment with you to see how good your memory is. I am going to show you a large card with a number of things fastened on it. You will have just half a minute to look at it. Half a minute is a pretty short time, so you must look very carefully, because afterwards I shall want you to tell me what you have seen, and I shall ask you questions about many little details, and I want you

[1] These objects are not exact duplicates of the Binet group, and the card is somewhat larger. The exposure-time and the questions of the interrogatory have been correspondingly modified.

to answer these questions exactly, if you can. Do you understand"?

Place the card directly before *S* in a good light. At the end of 30 sec., remove it and keep it well concealed. Direct *S* at once: "Now tell me everything you saw: describe it so clearly that if I had never seen the card I should know all about what was on it." The narrative is given orally by *S*, and recorded verbatim by *E*, without comment, query, or suggestion. Reread the report to *S*, and ask him to indicate what statements he is so sure of that he would swear to their accuracy. Underline these statements.

Proceed next with the interrogatory. If possible, ask *S* the following questions in the order given[1]. Record his replies by number, verbatim, and underline all attested replies.

Interrogatory for the card of objects.

(1) Did you notice a *coin?*
(2) What kind of a coin is it? (What denomination?)
(3) Does it show 'heads' or 'tails'?
(4) Is it bright or dull?
(5) Is it in good condition, or scratched and marred?

(6) Did you notice a *button?*
(7) What is its shape?
(8) What is its color?
(9) Is it the same color all over?
(10) Is it made of cloth or of some other substance?
(11) How many holes are there in it?
(12) How is it fastened to the cardboard?[2]

(13) Did you notice a *small picture* near the top of the cardboard?
(14) What color is it?
(15) What shape is it?
(16) What does it represent?
(17) How many persons are there in it?
(18) What is the lady doing with her right hand?
(19) What color is her dress?
(20) What is the other person doing?

---

[1] *S* may interfere with this program, either by anticipating the answers to some questions, or by committing errors, *e,g.*, describing an essentially different scene in the larger photograph; in such an event, *E* must devise other questions to follow up the cues thus given.

[2] If *S* replies "By a thread," ask further questions, *e g.*, "Do the threads pass through the holes or around the whole button?" "Draw them." "What color are they?" etc.

(21) Where is he sitting?
(22) What is he looking at? Describe it exactly.
(23) Is the name of the picture printed on it?
(24) Did you notice *another picture?*
(25) What shape is it?
(26) What color is it?
(27) What does it represent?
(28) How many persons are there in it?
(29) How are they dressed?
(30) Where are they standing?
(31) How many animals are there in the picture?
(32) Is the cart on wheels or not?
(33) Are there any words printed in the picture? What are they?
(34) What did you see in the background?
(35) What did you see in the foreground?
(36) Is the picture taken in summer or winter? How do you know?

(37) Did you notice a *stamp?*
(38) Is it American or foreign?
(39) How much is it worth? (What denomination?)
(40) What color is it?
(41) On what part of the cardboard is it?
(42) Is it a new one or has it been used? (Describe the cancellation-mark.)

(43) Did you notice a *label?*
(44) What color is it?
(45) What shape is it? (Is it perfectly rectangular? Draw it.)
(46) Is any printing on it? What?
(47) Is there any border around the printing?
(48) How is it fastened to the cardboard?
(49) How is it placed on the cardboard—right-side up, slanting, or how?
(50) What color is the cardboard?

VARIATIONS OF METHOD.—(1) Mature *S*'s may be tested in small groups, though this is not recommended. Both narrative and deposition must then be written by the *S*'s. For comparative purposes, the same procedure must be followed for all *S*'s, since oral and written reports cannot be assumed to be equivalent.

(2) To induce a moderate degree of suggestion, *E* may recast the questions of the above interrogatory into an expectative form, and add others, *e.g.*, in place of No. 12: "Is not the button fastened to the cardboard by a thread"? In place of No. 31: "Isn't there a little dog besides the horse"? In place of No. 42: "Isn't the

postage-stamp cancelled"? Or, for additions: "Isn't there a seventh object on the cardboard"? "Draw it." "Are there not four wheels on the cart"?, etc.

(3) To induce a strong degree of suggestion, E may recast the questions given into an implicative form, and add others as desired: *e.g.*, in place of No. 7: "Draw the button so as to show the place where it is broken." In place of 31: "Are both horses of the same color"? In place of 42: "Describe the cancellation-mark on the stamp." In addition to 45: "What else does the label have on it besides 'Glass. Handle with care.' "? Or, in place of 21: "Does the little boy's mother put her arm around him as he sits in her lap"? For additional questions, devise a number such as: "Is the lady's necktie dark brown or blue"?, etc.

RESULTS.—(1) In the narrative, Binet[1] found that, of 23 children, aged 9–12 years, only four mentioned all 6 objects, 10 mentioned 5 objects, 8 mentioned 4 objects, and 1 only 3 objects.

(2) In order of omission (for Binet's own objects), he found the stamp forgotten 10 times, the tag 9, the button 4, the coin 3, the portrait (smaller picture) 2 times, and the larger picture never.

(3) In tests of older children with written narratives, Binet found little difference in the total number of objects mentioned, but marked differences in the wealth of details and the precision of their formulation.

(4) The objects have distinct individuality, *i.e.*, though S may forget the color or the value of the stamp, yet if he recalls the object at all, it is as a stamp, not, for instance, as "some square, greenish-colored thing." In other words, S recalls a thing, not a number of meaningless attributes.[2]

(5) S's may report very precisely and with assurance objects or features of objects which are totally incorrect, *e.g.*, they may draw the thread fastening the button, and take oath as to its presence. Hence, testimony given with precision and detail and with the highest degree of assurance may be absolutely false.

(6) S's may recall one feature of an object exactly, but fail

[1] For a detailed presentation of these results, see his book, pp. 255-329.

[2] In the author's study of range of visual apprehension, however, there appeared numerous cases of the character thus denied by Binet, for example, a nickel was recalled only as "something bright and round in the upper corner of the cardboard."

entirely in their description of another feature of the same object *e.g.*, recall that the label is red, but err as to its shape. It follows that, in testimony, a witness whose assertions are verified in many details may, nevertheless, err in his statements with regard to some other detail that happens not to be susceptible of verification.

(7) If *S* fails to mention an object in his narrative, but recalls it immediately in the interrogatory, his further characterization of it may be quite as accurate as that of other *S*'s who had recalled it spontaneously.

(8) In comparing different types of questions, Binet found 26 per cent error for indifferent, 38 per cent for moderately suggestive, and 61 per cent error for strongly suggestive questions.

B. REPORT-TEST WITH A COLORED PICTURE

MATERIALS.—Set of four colored pictures: "Australians," "A Disputed Case," "Washington and Sally," and "The Orphan's Prayer."[1] Watch.

METHOD.—Give *S* instructions analogous to those in the preceding form of report-test, but without specifying the time of exposure. Expose the picture for 20 sec. Secure an oral narrative and deposition as directed above. Suggestions for interrogatories for two of the pictures follow.

Interrogatory for "A Disputed Case."

(1) How wide is the picture (horizontally)?
(2) How high is the picture (vertically)?
(3) Is there any border: if so, what color?
(4) How many persons are there in the picture?

Take the person on *your* right:

(5) Is he young, middle-aged, or old?
(6) What is his posture,—sitting, standing, or lying down?

[1] All four pictures may be procured through C. H. Stoelting Co., Chicago, Ill. The "Australians" is a large lithograph, one of a series called Leutemann's Types of Nations, catalogued by E. Steiger & Co., New York City. It is recommended for use with large groups, numbering from 10 to 50 or more *S*'s. The "Hindoos" lithograph prescribed in Test 31 may be used with this for check tests, as it is of the same dimensions and of similar character.

The "Disputed Case" (No. 1235 of the Taber-Prang Art Co's collection) is recommended for use save for very young children or for large groups. "Washington and Sally" and "The Orphan's Prayer" (Nos. 699 and 1207, respectively, of the same collection) may be used for subsidiary and check tests.

(7) What is he doing?
(8) What is his facial expression?
(9) Is he bald or has he abundant hair?
(10) What color is his hair?
(11) Is he smooth-faced or has he a moustache or a beard?
(12) What color is his beard?
(13) Does his moustache conceal his mouth?
(14) Does he wear eye-glasses or spectacles?
(15) Has he a hat on? What kind? What color?
(16) Where is his right hand?
(17) Where is his left hand?
(18) What color is his coat?
(19) What color is his shirt?
(20) Has he a collar on?
(21) What color is his necktie?
(22) What color is his vest?
(23) What color are his trousers?
(24) Does he wear slippers or shoes or boots?

Take the person on *your* left:

(25–44) Repeat questions 5–24.

(45) What kind of light or lamp is used?
(46) Where is it placed?
(47) Where is the ink-well?
(48) Is there not a pen in it?
(49) What color is the dog?
(50) Is there a table or bench?
(51) How long is it (really)?
(52) What color is the table cloth or covering?
(53) Is the fringe of the same or of a different color?
(54) Name the objects on the table.
(55) How many chairs are there in the room?
(56) Is the rocking chair on your left or on your right?
(57) Is there an umbrella?
(58) Do you think it is jet-black or dark-blue?
(59) In what position is it?
(60) Name the objects in front of the table on the floor.
(61) Is there a satchel or dress-suit case in the room? Which?
(62) Is it open or shut?
(63) What do the pictures on the wall represent?
(64) How many windows are visible?
(65) Can you see any detail of outdoor scenery through them?
(66) How many hats are there in the room?
(67) Describe and locate them.
(68) Can you recall the time indicated by the clock on the wall?

(69) What object is on your extreme right?
(70) Are there any books in this part of the room?
(71) What color is the wall?
(72) Where is the newspaper?
(73) How long did you see the picture?

Interrogatory for the "Australians."

(1) How many persons are there in the picture?
(2) How many animals?
(3) What kind of animals?
(4) What is the person on *your* left doing?
(5) What is the object behind him?
(6) What is the person in the center of the picture doing?
(7) Has this person a beard or not?
(8) Is the man who is in charge of the dog holding him by a leash (guiding rope) or by taking hold directly of the scruff of his neck?
(9) What are the persons in the background doing?
(10) Do the persons in the foreground wear anything beside the loin-cloth?
(11) What color is their skin?
(12) What color is the dog?
(13) What is the most peculiar thing that you noted in the appearance of the men in the picture?
(14) What objects lie in the immediate foreground?
(15) Is there any water represented in the picture?
(16) Is the white man standing on the left or on the right?
(17) Is the sun represented in the picture as shining from your right or from your left? How do you know?
(18) How long did you see the picture?

VARIATIONS OF METHOD.—Test the effect of varying the time of exposure, of extending the time-interval between exposure and report, of repeating the report (narrative or interrogatory), without further exposure, two or more times at intervals of several days or weeks.[1]

TYPICAL RESULTS.—The following narrative by a college senior, a man of varied experience, mature, much travelled, and well trained, though of mediocre native ability, shows clearly the tendency of an adult *S* to describe a situation, a meaningful whole, rather than merely to enumerate details, as do many children. Indeed, the detail here is distinctly subordinated to the interpre-

[1] See Ref. 17 for further suggestions.

tative rendering. The narrative tells what the picture is about. rather than what it is.

"The picture, about 10×10 inches, represents a scene that would be typical of a rural justice of the peace and a man who has come to ask his advice on some subject. The Justice sits before his desk, an old manuscript before him, one hand on his head as if he had not yet given his decision. The office is filled with books and on one of them in the left of the picture rests his top-hat. The visitor seems to be troubled very much: his clothing denotes that he is of a different station in life. He has placed his carpet-bag on the floor and his hat near it, as a sign of great mental strain, which seems to increase as he awaits the decision. On the wall to the right is a double map of the world, showing, perhaps, that the Justice is a man of wisdom and a source of information to his neighbors. The room, furniture, the manner of dress would have denoted a time long before ours. The men seem to be about 65 or 70 years of age."

In his deposition, this student rendered an unusually full list of answers: the reply—"I don't know"— is given only twice (Questions 34 and 72). The range of report is, therefore, large, but the fidelity is relatively small, since the following erroneous statements appear (those italicized are attested statements):

The picture is 14×14 inches. The man on the right *is bald, wears spectacles, has his right hand on a paper,* wears a collar, *a purple tie,* black trousers, and slippers. The man on the left is thinking hard, has a troubled expression, wears a sandy moustache: he has his right hand in his pocket, *his left on his knee:* he wears a light colored vest and brown trousers. The room is lighted by a candle which stands on the pile of books. There is a pen in the ink-well. The table is 14 feet long, has a light-colored cloth top *with fringe of a different color.* There are three chairs in the room, the rocker being at the left. *The umbrella is dark blue in color and lies on the floor. There is a coat on the floor in front of the table;* there is a basket on the table. *The satchel is shut. One window is visible.* There is a chair at the extreme right of the picture. The wall is white. (The cuspidor and the newspaper are not recalled.)

General Results of Tests of Report.—(1). *Accuracy.* The chief single result of the *Aussage* psychology is that an errorless report is not the rule, but the exception, even when the report is made by a competent *S* under favorable conditions. Thus, in 240 reports, Miss Borst found only 2 per cent errorless narratives and 0.5 per cent errorless depositions.[1]

[1] These errorless reports are commonly characterized by very small range: they are the reports of *S*'s who are extremely cautious and state only what they are certain of.

The average *S*, when no suggestive questions are employed, exhibits a *coefficient of accuracy* of approximately 75 per cent.

(2) *Range and accuracy.* There is no general relation of range to accuracy, though, for a given *S*, it is doubtless true that there is an inverse relation between these two coefficients.[1]

(3) *Range and other constants.* There is no general parallelism between range of report and other coefficients which depend upon degree of assurance.

TABLE 44

Comparative Accuracy of Sworn and Unsworn Statements (Stern and Borst)

EXPERIMENTER	STERN		STERN		STERN		BORST	
	Range	Errors	Range	Errors	Range	Errors	Range	Errors
Positive statements	(100)	13.6	(100)	19	(100)	23	(100)	11.0
Sworn statements	76	11	68	7	70	14	60	8.2
Unsworn statements	24	20	32	—	30	—	40	15.5
Certain statements							97.5	10.1
Uncertain statements							2.5	44.0

Note.—All figures are in per cents. The results, save those of the third and fourth columns, refer to narratives, not depositions.

(4) *Accuracy and attestation.* Generally speaking, attestation does not guarantee accuracy: on the contrary, though the number of errors is nearly twice as great in unsworn as in sworn testimony (according to Stern, 1.82 times, according to Borst, 1.89 times as great), there still remains as high as 10 per cent error in sworn testimony. These relations are shown clearly in Table 44.

(5) *Dependence on sex.* In all of Stern's work, both in narratives and depositions, with pictures, or events, or estimations of times and distances, whether under oath or not, the reports of men have been more accurate (by from 20 to 33 per cent), though less extended, than those of women, and a similar sex difference has

[1] The reason for this lack of general relation between range and accuracy is presumably that there are two kinds of good witnesses—the one possesses good capacity of observation, recall and report, and hence exhibits a large range and a high degree of accuracy; the other is cautious, and therefore restricts his range, which may be poor at best.

appeared in tests of school children. This superior accuracy of boys becomes more evident when the report is difficult to make. Stern's conclusions have, however, been criticized by both Wreschner and Miss Borst. Wreschner found that among adults women did better than men. Miss Borst likewise found women superior to men in accuracy and range, but inspection of her results shows that the superiority of women consisted in the fact that they returned a larger number of correct statements, and that the men did not make less accurate statements in their more limited reports.

More specifically, Borst found that in the narrative the range of men was 76 per cent, and in the deposition 83 per cent of the range of women, while the accuracy of men in both forms of report was approximately 96 per cent of the accuracy of women.

There is a similar discrepancy between Stern and Borst with regard to the tendency to attestation: the former found that men swore to 71 per cent, and women to 85 per cent of their report, whereas the latter found that men swore to 61 per cent, and women to but 59 per cent of their report.

(6) *Dependence on age.* The reports of children are in every way inferior to those of adults: the range is small, the inaccuracy large, and, since the assurance is high, the warranted assurance and reliability of assurance are both very low. During the ages 7 to 18 years, the range, especially the range of knowledge, increases as much as 50 per cent, but the accuracy, save in the deposition, does not increase as rapidly (20 per cent). This development of capacity to report is not continuous, but is characterized by rapid modification at the age of puberty.

The one factor that more than any other is responsible for the poor reports of children is their excessive suggestibility, especially in the years before puberty.[1]

Stern has endeavored to analyze in part the development of the child's capacity to report, and has distinguished four stages: (1) the very young child enumerates only isolated objects or persons (Binet's enumerator type); (2) at about the eighth year, actions are reported more carefully; (3) during the years 9–10, attention is for the first time paid to spatial, temporal, and causal relations; (4) in a still later period, there appears the capacity

[1] On the general subject of children's reports, consult the work of Binet, Stern, Lobsien, Borst, and Plüschke.

to make a qualitative analysis of the constituent features of the objects reported.

(7) *Dependence on intelligence.* We have as yet no conclusive experiments upon the relation between accuracy of report and general intelligence.

(8) *Defectives.* The reports of defectives, paralytics, epileptics, the insane, etc., show, as one might expect, a very high degree of inaccuracy, even when the pathological condition is not seriously developed. Such persons are also highly suggestible (de Placzek).

(9) *Dependence on time-interval.* Lengthening of the time-interval between experience and report exerts, as one might expect, a generally unfavorable influence, but there is nothing like the loss in efficiency shown in curves of memory for nonsense syllables, as in the familiar tests of Ebbinghaus: indeed, for some *S*'s the report seems to be somewhat improved after several days have elapsed, and, in general, the conditions are so complex as to demand further special investigation.

TABLE 45

Effect of Time-Interval on Range and Accuracy of Report (Borst)

FORM OF REPORT	NARRATIVE		DEPOSITION	
Intervals in Days......	3	9	3	9
			Per cent	*Per cent*
Range..................	40.6	39.6	67.2	65.5
Accuracy...............	89.5%	87.9%	82.6	83.4

From his earlier tests, Stern computed a fairly constant decrease of accuracy with time, amounting, on the average, to a loss of 0.33 per cent per day over the period of three weeks which he studied; similarly, Borst computed a decrease in accuracy of 0.27 per cent per day during a period of six days. Some of her results are presented in Table 46.

Though range and accuracy seem thus to suffer with the lapse of time, assurance, as shown by the number of certain and attested statements, is not, it seems, equally affected, but shows either a surprising constancy, or, if anything, a tendency to increase. From this it may be concluded that assurance and tendency to oath are due to *S*'s 'personal equation,' rather than to the freshness of his memory. It would follow, of course, that warranted assurance and warranted tendency to oath decline with the lapse of time.

(10) *Dependence on contents or features.* Not all the features of the original experience are reported with the same frequency or with the same accuracy: there is, rather, a process of selection, both in the process of observation, and also, probably, in memory and in the formulation of the report. In general, we may say that persons and their acts, objects, things, and spatial relations are reported with considerable accuracy (85–90 per cent), whereas secondary features, especially quantities and colors, are reported with considerable inaccuracy (reports on color have an error of from 40 to 50 per cent).

(11) *Dependence on form of report.* All authorities agree that the use of the interrogatory, whether of the complete or incomplete form, increases the range and decreases the accuracy of the report. Thus, in comparison with the narrative, the range of the interrogatory may be 50 per cent greater, while the inaccuracy (of the incomplete interrogatory) may be as much as 550 per cent greater. In general terms we may say that about one-tenth of the narrative is inexact, but about one-quarter of the deposition. Typical statistics are given in Table 46.

TABLE 46

Dependence of Report on its Form (Stern and Borst)

AUTHOR	RANGE		ACCURACY	
	Narrative	Deposition	Narrative	Deposition
			Per cent	*Per cent*
Stern....................	25.5	52.1	94	67.1
Borst....................	40.5	65.6	89	83.0

Note.—In comparing these figures, it should be remembered that Stern used an incomplete, and Borst a complete interrogatory.

(12) *Dependence on the type of question.* The introduction of leading or suggestive questions very noticeably decreases the accuracy of report for children, and, unless the conditions of report are quite favorable, even for adults. The greater suggestibility of children is shown by Stern's results in which the inaccuracy of boys and girls aged 7 to 14 was from 32 to 39 per cent, as against 10 per cent inaccuracy for young men aged 16 to 19 years. Binet's results with suggestive questions have already been cited.

(13) *Dependence on the ideational type of the reporter.* The best reports are given by observers of a mixed ideational type, *e.g.*, acoustic-motor or visual-motor (Borst): even in a picture-test, the purely visual-minded observer is inferior, though less open to suggestion (Lobsien).

A characteristic analysis of reports, for the purpose of classifying reporters into ideational types has been given in the description-of-an-object test (No. 31), in which Binet distinguishes four types of reporter—the observer, the describer, the emotionally-minded, and the erudite. Miss Borst was unable to use this classification, however, with her *S*'s.

Another classification of reporters according to mental type was attempted by Miss Borst, who, after a preliminary tachistoscopic test, compared the reports of 'fixating' and 'fluctuating' *S*'s, and concluded that *S*'s whose attention is of the 'fixating' type have uniformly the greater warranted assurance of report.

(14) *The effect of repeating a report.* When *S* is called upon to make his report several times, the effect of this repetition is complex, for (1) it tends in part to establish in mind the items reported, whether they be true or false, and (2) it tends also to induce some departure in the later reports, because these are based more upon the memory of the verbal statements of the earlier reports than upon the original experience itself, *i.e.*, the later reports undergo distortion on account of the flexibility of verbal expression.

(15) *The effect of practise.* Simple practise in reporting, even without special training or conscious effort to improve, facilitates and betters the report, as is shown in Table 47, from Miss Borst. It will be noted that the tendency to oath and warranted tendency to oath are both particularly improved by practise, and that there is also an appreciable improvement in range, accuracy, warranted assurance, and reliability of assurance, whereas assurance and accuracy of assurance are scarcely affected. Similar practise-effects may be discerned in the deposition. From these results, it is clear that the several coefficients of report may vary more or less independently.[1]

[1] There are needed, in the author's opinion, further experiments with careful introspective control by trained adults, so that we may secure a more adequate analysis of the factors that are affected by practise: the previous investigators have relied too much upon mere statistical evaluation, whereas the field is now ripe for an introspective analysis of the mental operations concerned in report—an analysis that should, of course, make the utmost use of the statistical method for check and control.

TABLE 47

Effect of Practise upon Coefficients of Report (Narrative) (Borst)

NUMBER OF REPORT (TEST)	I	II	III	IV	V
Range	39.0	39.0	42.3	40.3	42.0
Accuracy	86.6	87.7	92.9	88.2	90.0
Assurance	96.6	96.4	97.8	97.9	98.6
Warranted assurance	84.0	87.0	91.0	88.0	89.0
Reliability of assurance	87.5	89.4	92.6	89.8	90.3
Accuracy of assurance	97.0	98.0	98.4	98.6	99.2
Tendency to oath	43.0	59.8	62.8	61.9	72.1
Warranted tendency to oath	40.2	53.2	58.5	57.5	66.5
Unwarranted tendency to oath	2.8	6.6	4.3	4.4	5.6
Reliability of oath	93.0	88.8	92.5	93.0	91.7

Note.—The effect of practise in these tests is somewhat obscured by the fact that the first and third tests were made after a 3-day, the others after a 9-day interval.

(16) *The educability of report.* The capacity of children to observe and report in a detailed and accurate manner may be improved by systematic training. This education may be best secured by appeal to zeal, interest, enthusiasm, or desire for improvement on the part of the child; more formal training of an intellectual type—*e.g.*, suggestions for systematic observation, specific training in sense-perception, instruction designed to augment appropriate apperceptive-masses, etc.—is much less effective.

The inadequacy of the child's report is due, not so much to poor memory, as to the fact that he fails to perceive many features in the original experience, that he fails to put into words even what he does perceive, and especially to the fact that he is absurdly uncritical (his assurance, indeed, commonly reaches 100 per cent).

REFERENCES

A. The most important single source is Stern's *Beiträge zur Psychologie der Aussage*, Leipzig, 1903–6. Lack of space forbids the itemizing of the numerous titles: besides extended reviews, communications, reports of lectures, etc., this periodical contains important articles by Stern, Jaffa, Cramer, Lobsien, Lipmann, Borst, Bogdanoff, Rodenwaldt, Oppenheim Kosog, Wendriner, Günther, Gottschalk, and others.

B. The following are other important references. See No. 17 and especially No. 18 for bibliographies.

(1) A. Binet, La Suggestibilité, Paris, 1900. Pp. 391.

(2) A. Binet, La science du témoignage, in A. P., 11: 1904 (1905), 128–137.

(3) A. Binet, Psychologie individuelle. La description d'un objet, in A. P., 3: [1896] (1897), 296–332.

(4) Marie Borst, Recherches exp'tlles sur l'éducabilité et la fidélité du témoignage, in Ar. P., 3: 1904, 233-314.

(5) Marie Dürr-Borst, Die Erziehung der Aussage und Anschauung des Schulkindes, in E. P., 3: 1906, 1–30.

(6) M. Borst and E. Claparède, La fidélité et l'éducabilité du témoignage, in Arch. des sciences physiques et naturelles, April 7, 1904.

(7) De Placzek, Exp. Untersuchungen über d. Zeugenaussagen Schwachsinniger, in Gross' Archiv f. Kriminalanthropologie, 18: 1904, 22-63.

(8) O. Lipmann, Neuere Arbeiten zur Psychologie der Aussage: Sammelbericht, in Journ. f. Psych. u. Neurologie, 3: 1904, 245–249.

(9) O. Lipmann, Die Wirkung der Suggestivfragen, in Z. P. P., 8: 1906, 89–96.

(10) O. Lipmann, Die Wirkung v. Suggestivfragen, in Z. A. P., 1: 1907-8, 44-92, 382–415, 504–546.

(11) O. Lipmann, Praktische Ergebnisse der Aussageforschung, in Z. P. P., 97–103.

(12) M. Lobsien, Ueber Psych. der Aussage, in Z. P. P., 6: 1904, 161–209.

(13) L. Maurer, Beobachtungen über das Anschauungsvermögen der Kinder, in Z. P. P., 5: 1903, 62–85.

(14). H. Münsterberg, On the Witness Stand, N. Y., 1908. Pp. 265. (The several sections of this book have also appeared in magazine form, chiefly in McClure's Magazine.)

(15) L. W. Stern, Zur Psych. d. Aussage. Exp. Untersuchungen über Erinnerungstreue, in Zeits. f. d. ges. Strafrechtwissenschaft, 22: 1902 (also published separately, Berlin, 1902).

(16) L. W. Stern, Literatur zur Psych. d. Aussage, in Z. A. P., 1: 1907–8, 429-450.

(17) G. M. Whipple, The observer as reporter: a survey of the 'psychology of testimony,' in P. B., 6: May 15, 1909, 153-170.

(18) J. H. Wigmore, Professor Münsterberg and the psychology of evidence, in Illinois Law Review, 3: Feb., 1909, 399-445 (with bibliography of 127 titles).

(19) A. Wreschner, Zur Psych. d. Aussage, in A. G. P., 1: 1905, 148–183.

CHAPTER IX

Tests of Association, Learning, and Memory

A generation ago, the members of the 'English School' of psychologists exalted 'association' as a fundamental principle or law of mind comparable in its scope and importance with the law of gravitation in the material world. Whether this extreme position be held or not, it must be admitted that the more complex phases of mental activity are more readily understood if certain basic principles of mental elaboration are posited, particularly the principles: attention, retention, and association. Disregarding the first of these, which we have already discussed, we find in retention the *sine qua non* of the development of human mental activity, and we find constantly at work in the conscious life of the organism a tendency for the establishment of connections between its concurrent and its successive psychophysical activities. In so far as the conscious organism acquires new capacities for response, there must be retention and organization. Learning, retaining, recalling, associating, these are terms obviously descriptive of a series of related activities, and on this account, tests which deal with them are here assembled.

The first three tests in this chapter investigate the nature and the efficiency of those associative connections that the subject has already established at the time of the experiment, whether the associative processes are allowed free rein, or are placed under certain restrictions. The two learning tests investigate the subject's capacity to establish new associative connections, under relatively novel conditions. The memory tests, in a somewhat different way, investigate his retentive capacity or his ability to reproduce an arbitrary series of symbols or a series of related ideas.

Association and memory, taken together, have undoubtedly been the occasion of more numerous and more elaborate experimental investigations than any other phase of mental life. Learn-

ing, in the narrower sense, has, perhaps, received somewhat less attention, though of late the importance of its application to pedagogical problems has stimulated work upon it.

The experimental study of associative activity can be, and has been, undertaken for quite varied purposes, *e. g.*, to examine the time relations of mental phenomena, to study individual differences in thought-processes, as conditioned by age, sex, training, physical condition, and the like, to analyze the diurnal curve of psychophysical efficiency (as in Kraepelin's use of computation), to diagnose mental content, and even to reveal obscure mental tendencies and motives or intentionally withheld information (diagnostic association tests). Space forbids the exploitation of all the tests that have been developed in these fields, but a study of the more common tests of learning, association, and memory that have been selected for treatment here as being most applicable to the experimental study of school children, will serve to indicate the lines along which variant methods may be developed and employed.

TEST 33

Uncontrolled association (continuous method).—The essence of this test is the requirement to write or pronounce an extended series of words not in the form of sentences. Its interest lies, first, in the difference of facility exhibited by different *S*'s in the production of such a series of terms; secondly, in the nature of the terms given by *S*'s of different sex, age, or social condition; and thirdly, in the nature of the mental processes underlying the word-naming process.

Cattell and Bryant (3) make brief mention of the test; Jastrow (5, 6), and later Miss Nevers (7) Miss Calkins (2), and Miss Tanner (8), employed it for the study of the community of ideas of men and women, Flournoy (4) for the study of the effect of environment, present and immediately past, upon the course of association, and Binet(1) for the study of individual differences in intellectual processes.

Materials.—Stop-watch. Blank forms containing numbered spaces for 100 words.

Method.—Give *S* these instructions: "When I say 'now,' I

want you to start in with some word, any one you like, and keep on saying words as fast as you can until you have given a hundred different words. You may give any words you like, but they must not be in sentences. I will tell you when to stop." *E* starts the stop-watch at the command 'now' and writes on the prepared form the words spoken by *S*. With mature *S*'s, it may be possible to get nothing more than scant abbreviations for the more rapid portions of the series, but these may be filled out subsequently. At the 100th word, stop the watch and record the time. If the time permits, and *S* can do so, it is advisable at once to go over his series and make marginal notes of all the intermediate links and subsidiary associative processes that he can recall.

Variations of Method.—(1) For group tests, *E* may provide each *S* with a blank containing 100 numbered spaces. Allow 3 min. for writing, and rate speed in terms of number of words written.

(2) *E* may omit the instruction to write or to speak as rapidly as possible, and allow *S* to work at his leisure. This method, which was followed by Miss Nevers, is perhaps more satisfactory for the subsequent qualitative report upon the series, but deprives the test of whatever quantitative merits it possesses, besides tending to yield results of a distinctly different nature that are not comparable with those otherwise obtained.

(3) When working with younger *S*'s, *E* may with advantage limit the length of the series. Thus, Flournoy demanded but 10 words, while Binet recorded the time for three series of 20 words each, and occupied the intervals in reviewing with *S* the terms of the preceding series. This method is less fatiguing, and enables immature *S*'s to give a more satisfactory account of their associative connections, but it does not test *S*'s capacity as rigorously as the longer list.

(4) *E* may secure a very limited measure of uniformity in the earlier portion of the series by starting all *S*'s from the same word. For this, the words *quick* and *play* are recommended. Here it is of interest to observe the lines of divergence in association taken by different *S*'s.

(5) Another variation is that of Flournoy, who, in addition to the word test, gave 45 *S*'s instructions to make 10 drawings of any sort.

TREATMENT OF DATA.—In the standard form of test, *S*'s speed is indicated directly by his time for naming 100 words. In the group test, it is customary, similarly, to rate *S*'s speed in terms of words written in 3 min. It is not possible, however, to regard the times obtained from these two forms of the test as interchangeable, since the second form includes writing and this, as is demonstrated below, tends, even in the case of mature *S*'s, to slow the rate of performance. In so far, too, as *S*'s differ in their speed of writing, this fact enters as an unavoidable disturbing factor in the group test.

For qualitative comparison of the lists, *E* may, by inspection, supplemented by *S*'s explanations, catalog the words, either in the 7 categories used by Binet, or in the 25 categories used by Jastrow and Miss Nevers. Both classifications are embodied in the results below.

RESULTS.—(1) In tests of college students, Jastrow found an *average time* of 130 sec. for oral, and 308 sec. for written lists of 100 words. Since writing an equal number of words from dictation took 212 sec., Jastrow concludes that, in naming 100 words, about 1.14 sec. is used, on the average, in thinking the association between one word and the next.

(2) Inspection of the lists printed both by Jastrow and by Binet shows that *S*'s follow what might be termed a *series of themes*: a number of terms are written, all of which cluster about a common central idea; through one of these terms access is given to a new central idea, which in turn becomes a theme for the next series of terms; thus, in the series *hand, face, lip, chest, knees, calf, cow, horse, pig,* etc., the transition from the parts-of-the-body theme to the animal theme is effected by the common term *calf.*

(3) In some *S*'s, the controlling theme is an *auditory sequence,* which occasions long series of rimed or alliterative terms, *e.g.,* *run, pun, fun,* etc., or *hen, hand, head, harp,* etc.

(4) In this test, the *most common words, i. e.,* those most easily got at, or those that lie, as it were, on the surface, are given first. After these are delivered, the task grows more difficult; deeper and more remote-lying terms must be actively sought for. Closely related to this is the fact that, at least in the lists of younger *S*'s, practically all the terms are nouns. This is par-

ticularly the case in the short series conducted by Binet, so that, as he remarks, the test, as he conducted it, is virtually equivalent to a request to write 20 common nouns.

(5) In view of the vast number of words available, it is at first surprising to note the *degree of community* present in lists of 100 terms given by a limited number of persons. Thus, Jastrow found that in 50 lists (5000 words), only 2024 words were different, only 1266 words occurred but once, while the 100 most frequent words made up three-tenths of the whole number.

These most frequent words are, as has just been said, names of common objects: in Jastrow's 50 lists, the following were the most frequently used words: *book* (40), *horse* (37), *girl* (35), *man* (34), *boy* (33), *table* (30); then follow *chair, tree, cow, paper, dress,* etc., in somewhat lesser frequency.

(6) For the *classification* of the words given by 20 12-year old pupils, Binet found seven categories adequate, viz: (*a*) names of objects in the room where the test was held, (*b*) parts of the person or clothes, (*c*) objects or persons in the school, (*d*) objects recalled from the home, (*e*) objects seen in the streets (horse, tree), (*f*) objects seen in fields or on country excursions, (*g*) unclassified nouns. Here there is no place for abstract terms, many of which were found in series given by American pupils in Jastrow's tests. Jastrow's own classification is indicated in Table 48, where it will be seen that his 25 categories are much more elaborate and extended than those employed by Binet.

(7) The question as to *sex difference* in spontaneous trains of ideas such as are evoked in this test has been answered differently by the tests conducted at Wisconsin University and at Wellesley College. The comparison of Wisconsin men and Wisconsin women was made by Jastrow, the 1894 test of Wellesley women by Miss Nevers and with no instruction as to speed, the 1896 test of Wellesley women by Miss Calkins but with the same instructions as those of Jastrow. The categories of particular interest are those printed in italics. Jastrow's results in this and other tests led him to believe that "the women repeat one anothers' words much more than the men." He found that "the class to which women contribute most largely is that of articles of dress, one word in every eleven belonging to this class. The inference from this that dress is the

TABLE 48

Distribution of Terms in 'Uncontrolled' Association (Jastrow, Nevers, Calkins)
(Each column represents 25 lists of 100 words each.)

CATEGORIES	WISCONSIN MEN	WISCONSIN WOMEN	WELLESLEY WOMEN, 1896	WELLESLEY WOMEN, 1894
1. *Animal kingdom*	254	178	146	223
2. *Wearing apparel and fabrics*	129	224	97	96
3. Proper names	194	153	81	141
4. Verbs	197	134	279	114
5. Implements and utensils	169	121	139	132
6. *Interior furnishings*	89	190	212	84
7. Adjectives	177	102	300	234
8. *Foods*	53	179	88	56
9. Vegetable kingdom	121	110	101	91
10. *Abstract terms*	131	97	101	280
11. Buildings and building materials	105	117	86	106
12. Parts of body	101	105	66	34
13. Miscellaneous	91	97	123	162
14. Geographical and landscape features	97	80	70	142
15. Mineral kingdom	74	96	30	54
16. Meteorological and astronomical	85	76	109	26
17. Stationery	60	86	69	26
18. Occupations and callings	71	47	24	33
19. Conveyances	62	52	19	79
20. *Educational*	34	76	102	167
21. Other parts of speech	96	5	164	41
22. *Arts*	33	61	17	44
23. *Amusements*	30	53	17	102
24. Mercantile terms	30	29	18	15
25. Kinship	17	32	42	18

predominant category of the feminine (or of the privy feminine) mind is valid with proper reservations." Since the women exceed the men in the enumeration also of foods, amusements, arts, and

educational matters, but fall below them in naming implements and utensils, professions, and especially in abstract terms, Jastrow concludes, "that the feminine traits revealed in this study are an attention to the immediate surroundings, to the finished product, to the ornamental, the individual, and the concrete, while the masculine preference is for the more remote, the constructive, the useful, the general, and the abstract" (5; pp. 564–5). Most of these conclusions are flatly opposed by the Wellesley results of 1894, but the employment of identical methods in the 1896 test produced less marked divergencies. It is particularly to be noted that writing at a faster rate (1896 test) caused a marked decrease in the number of abstract terms, and brought the terms relating to 'interior furnishings' up even beyond those of the Wisconsin women; on the other hand the frequency of terms for 'wearing apparel' was not affected by this change in method.

These discrepancies raise the issue, as Miss Tanner has pointed out, whether this test can be expected to reveal fundamental native differences in mental constitution of the two sexes, or whether it reveals merely acquired traits, social traditions, individual habits, educational, and other environmental influences. The lists written by college students might be expected, for example, to be considerably affected by their recent occupations, courses of study pursued at the time, etc.

(8) This *influence of environment* upon the lists of associations is indicated particularly in Flournoy's brief tests (10 words and 10 drawings), the results of which are summarized in Table 49.

TABLE 49

Influences that Affect 'Uncontrolled' Series of Words or Drawings (Flournoy)

	DRAWINGS	WORDS
	per cent	*per cent*
Traced to present surroundings	13.8	29.0
Traced to the immediate past	1.9	8.2
Due to the *milieu*	15.7	37.2
Traced to recent personal experiences	2.4	3.9
Traced to personal habits	39.2	9.2
Expressing individuality	41.6	13.1
Unexplained	42.7	49.7

REFERENCES

(1) A.Binet, L'Étude expérimentale de l'intelligence, Paris, 1903. Pp. 309. Especially chs. ii. to iv.

(2) Mary W. Calkins, Community of ideas of men and women, in P. R., 3: 1896, 426-430.

(3) J. McK. Cattell and Sophie Bryant, Mental association investigated by experiment, in Mind, 14: 1889, 230-250.

(4) Th. Flournoy, De l'action du milieu sur l'ideation, in A. P., 1: 1894 (1895), 180-190.

(5) J. Jastrow, A study in mental statistics, in the New Review, 5: 1891, 559-568.

(6) J. Jastrow, Community of ideas of men and women, in P. R., 3: 1896, 68-71, 430-1.

(7) Cordelia Nevers, Dr. Jastrow on community of ideas of men and women, in P.R., 2: 1895, 363-7.

(8) Amy Tanner, The community of ideas of men and women, in P. R., 3:, 1896, 548-550.

TEST 34

Controlled association: part-wholes, genus-species, and opposites.—The common features of these tests, which are characterized by their users as tests of "ability to appreciate relationships and to control associations," is the demand for the execution of a verbal association which is so restricted that only a very limited number of words may be considered as correct associates. The restriction is less stringent in the part-wholes and genus-species than in the opposites test. On account of this restriction of connection, the tests may be said to occupy a ground intermediate between the uncontrolled series illustrated in the preceding test, and the strictly controlled associations involved in the computation tests that follow. The three tests have all been elaborated at Columbia University. Examples of their use may be found in Miss Norsworthy's tests of feeble-minded children (2), in Thorndike's study of the degree of resemblance of twins (5), and in the examination of correlations of perceptive and associative processes by Aikins, Thorndike, and Hubbell (1).

A. THE PART-WHOLES TEST

MATERIALS.—Stop-watch. Printed form containing the ten nouns mentioned below, and provided with spaces for the recording of an equal number of associates.

METHOD.—Instruct S as follows: "I shall give you a paper on which are printed ten words. I want you, as rapidly as you can, to give for each word the name of the whole thing of which the word is a part. For instance, if the word *fur* were given, you would be asked to name a thing that *fur* is a part of, like *cat*; for *hand* you might say *arm* or *watch*. To make sure you understand, give me a 'whole' for *mercury*—for *drawer*." If S displays proper comprehension of the test, give him the printed list of test-words; start the stop-watch as he glances at the first word; write down his associates on a separate blank as fast as given, and record the total time.

VARIATIONS OF METHOD.—(1) For a group test, each S is supplied with the printed blank and is requested himself to write the associates after each test-word. To retain the measure of speed as well as accuracy of association, the test must be conducted with a time-limit such that the fastest S in the groups to be tested can but just finish the ten words. This unavoidable introduction of writing necessarily makes the quantitative results differ from those obtained by oral report.

(2) Following the procedure adopted by Miss Norsworthy, the group-test may be given with no time-limit: here efficiency is figured entirely on the basis of qualitative performance.

TREATMENT OF DATA.—(1) When time is recorded, speed is measured directly in seconds, or in words, according as the test is given to individuals or to groups. It would, perhaps, be possible to combine speed and accuracy in a single index of efficiency after the methods described in the Cancellation Test (No. 26.) Suggestions for other methods of computing net efficiency are given in the discussion of the opposites test.

(2) The method of scoring used by Miss Norsworthy is, in the author's opinion, somewhat open to criticism, but it is reproduced here in order that E's may compare their results with her published norms. Her scores are based simply upon the number of associates correctly given. A perfect score is, therefore, 10. The key employed is as follows:

WORD	ASSOCIATES DEEMED CORRECT.	ASSOCIATES DEEMED INCORRECT.
Door	Anything that usually has a door	
Pillow	Couch, bed	Sofa
Letter	Word, alphabet, envelope	
Leaf	Tree, plant, book	
Button	Anything usually having buttons	
Nose	Face, head	Cheek
Cover	Book, bed, kettle	
Page	Book	
Engine	Train, car	
Glass	Window, door	Tumbler

This key is, of course, not exhaustive, but simply illustrative.

Results.—(1) Table 50 represents the performance of 504 *normal children,* aged 8 years and over, as reported by Miss Norsworthy.[1] Sex-differences are not sufficiently evident to justify separate treatment.

TABLE 50

Normal Performance in the Part-Wholes Test (Norsworthy)

AGE	8	9	10	11	12	13	14	15	16	ADULTS
Median	6.5	7.8	7.8	8.7	8.7	9.0	9.0	9.0	9.0	10.0
P. E.	2.3	1.3	1.9	1.1	1.2	0.7	0.7	0.7	0.7	0.5

(2) *Defective children,* according to the same writer, are distinctly inferior to normal children in this test: thus the percentage of normal children with a record above the median, above −1 P.E., and above −2 P.E., would of course, be 50, 75, and 91, respectively, but the percentages of feeble-minded children obtaining these three grades of efficiency were but 9, 17, and 27, respectively. That is, only 9 per cent of the feeble-minded children reached the degree of efficiency attained by one-half of the normal children, etc.

B. THE GENUS-SPECIES TEST

Materials.—As in the part-wholes test, but with a different list.

[1] These norms are perhaps unduly high in some cases, since three of the ten test-words were actually given as examples before the test was administered.

METHOD.—Instruct S as follows: "I shall give you a paper on which are printed ten words. I want you, as rapidly as you can, to give for each word the name of some particular thing, the class name of which is given, *i.e.*, you are to give some species of the genus meant by the given word. For instance, if the word *animal* were given, you would be expected to name some one kind of animal, like *horse*, or if the word *verb* were given, you might give *run* as a particular verb. To make sure you understand, give me a word for something in the class named by the word *man*. Give me a 'species-associate' for *money*." When S is clear as to what is wanted, the test is administered in the same manner as the part-wholes test.

VARIATIONS OF METHOD.—The same variations may be adopted as in the part-wholes test.

TREATMENT OF DATA.—(1) Speed or rate of performance is computed as in the part-wholes test.

(2) The method of scoring adopted by Miss Norsworthy, which is the one on which the results below are based, takes account only of the number of correct associates written. In general, any specific term that falls under the class term is counted as correct. The word *play* was not allowed under *toy*, nor the word *toy* under *game*.

RESULTS.—(1) Table 51 represents the *performance of normal children*, based on the examination of 511 children of both sexes.

TABLE 51

Normal Performance in the Genus-Species Test (Norsworthy)

AGE	8	9	10	11	12	13	14	15	16	ADULTS
Median	5.0	5.0	7.0	9.2	9.2	9.3	9.3	9.5	9.5	10.0
P. E.	2.0	2.7	2.9	1.9	0.7	0.4	0.5	0.5	0.5	0.0

(2) *Feeble-minded children* are distinctly inferior in this test to normal children of the same age: only 9 per cent reach the median of the normal children, only 16 per cent reach − 1 P.E., only 17 per cent reach − 2 P.E. of normal children of their age.

C. THE OPPOSITES TEST

MATERIALS.—Stop-watch. Printed forms embodying one or all of the following lists, with blank spaces for the recording of associates.

List A is No. 1 and List B is No. 2 as used by Miss Norsworthy for testing on two successive days. List B is also the one mentioned by Thorndike in his study of twins and elsewhere, also by Aikins, Thorndike and Hubbell. The last-named investigators used List C for their so-called 'hard-opposites test.' These three lists have the advantage of extended use, so that norms can be stated for them with some confidence. Wells, however, contends that they contain unequally difficult stimulus-words, and proposes several other lists, one of which is as follows: *High, heavy, up, worse, few, true, east, wrong, wet, asleep;* another ten-word list suggested by Wells is: *after, sick, smooth, early, large, open, good, weak, long, glad.*

LIST A.	LIST B.	LIST C.
bad	good	stupid
inside	outside	hard-working
slow	quick	strong
short	tall	sane
little	big	obnoxious
soft	loud	foolish
black	white	handsome
dark	light	adroit
sad	happy	superior
true	false	loquacious
dislike	like	rapid
poor	rich	generous
well	sick	straight
sorry	glad	separate
thick	thin	up
full	empty	always
peace	war	joy
few	many	high
below	above	obscure
enemy	friend	proud

METHOD.—Instruct *S* as follows: "I shall give you a paper on which are printed 20 words. I want you, as rapidly as you can, to give for each word, a word that means just the opposite to it, that is, a word that means just what the printed word doesn't mean. For instance, the opposite to *day* is *night,* the opposite to *front* is *back.* To make sure you understand, give me an opposite for *old*—

for *sharp*." If S has difficulty, E may seek other means for making the nature of the test clear, *e.g.*, "If a man isn't *sick*, what s he?" "One road is *long*, another might be what"?, etc. The test is then administered in the same manner as the part-wholes test.

VARIATIONS OF METHOD.—The same variations may be adopted as in the part-wholes test. To test the effect of practise, a reverse series may be given immediately after the test or, say, on the following day. List B, it is evident, is a reverse of List A, and was used by Norsworthy in the manner just mentioned.

The last-named investigator allowed 60 sec. for the opposites test in her study of feeble-minded children, after three of the words on the list had had their opposites given for illustration. Thorndike also gave 60 sec. for the test, but Aikins allowed only 30 sec. for either List B or List C. It is to be remembered that in all these cases the associates were written. The better method is that prescribed: oral response by individual S's, timed by E.

TREATMENT OF DATA.—Various plans have been followed in scoring this test, some of which seem needlessly complicated and artificial. The simplest plan is rank S's in terms of speed (time needed or words written in a limited time) and in terms of accuracy (percentage of correct opposites). If time is neglected, S's may be scored by simply subtracting the number of wrong associates from the number of right associates, neglecting possible instances in which no associate is given.[1]

A modification of this simple method of scoring which will admit of giving partial credit for some associates that are not strictly correct, yet not wholly wrong, is illustrated in the work of Miss Norsworthy, who allowed a credit of 1 for each correctly given associate and 0.5 for partially correct associates. The correct associates for List A are the terms of List B, and *vice versa*. In addition to these words, certain other terms are allowed either full credit or half credit, as indicated herewith.

STIMULUS	ASSOCIATES CREDITED WITH 1.	ASSOCIATES CREDITED WITH 0.5
inside		out, outdoors
slow	fast	quickly
short		big

[1] This is the method of scoring represented in Thorndike's curve of distribution for this test (4: p. 49), and used by the same author in his study of mental resemblances in twins.

STIMULUS	ASSOCIATES CREDITED WITH 1.	ASSOCIATES CREDITED WITH 0.5
little		tall
soft	hard	rough
dark	daylight	
sad	glad	
true		falsehood
dislike	love	
well	ill, badly	
sorry	happy	
few	a lot	
below	over, on top	
enemy	companion	
outside		in, indoors
quick		lazy, slowly
tall		little, low
big	short	
loud	low	whisper
happy	sorry	sorrow
false		right, truth
like	unlike, disliked, different, hate	
sick	healthy	
glad	mad	
thin		broad
empty	filled	
many	one	
above	under	

In the limited-time method adopted by Aikins (30 sec. per list), scoring was effected in terms of five quantities: (1) the point reached in the test, (2) the number of associates written, (3) the number of errors, (4 the percentage of errors to the number done, and (5) a general index (net efficiency), calculated in the hard-opposites test (List C) by subtracting twice the number of errors from the number known, multiplying the remainder by 2, and subtracting the number of omissions, and in the easy-opposites test in the same manner, save that 3 times the number of errors were subtracted.

RESULTS.—(1) The *performance of normal children* for Lists A and B, as scored by Miss Norsworthy, is indicated in Table 52.

(2) *Sex differences* cannot be made out with certainty.[1]

(3) In comparing *feeble-minded* with normal children, the former are found markedly inferior. No feeble-minded child, according

[1] See also Thorndike (3, p. 117). For the distribution of 239 cases of 12.5 year old children, see the same author (4, Fig. 32, p. 49).

TABLE 52

Normal Performance in the Opposites Test. Both Sexes. 605-608 Cases (Norsworthy)

AGE	FIRST LIST		SECOND LIST	
	Median	P. E.	Median	P. E.
8.0	7.4	2.0	8.7	1.4
9.0	9.0	2.0	9.5	1.7
10.0	9.9	3.0	11.5	2.2
11.0	12.5	3.0	13.1	2.9
12.0	13.5	2.6	14.7	3.6
13.0	14.0	2.5	16.4	2.4
14.0	14.5	2.3	17.8	2.0
15.0	15.0	2.3	18.5	2.0
16.0	15.5	2.3	19.0	2.0
Adults	20.0	1.0	20.0	1.0

to Miss Norsworthy's tests, reached the level of the median performance of normal children of his age, only about 1 per cent were better than − 1 P.E., and only 6 per cent better than − 2 P.E. of normal children of their age.

(4) There is a high degree of correlation ($r = 0.90$, P.E. about .05) between the *capacity of twins*, as measured by Thorndike.

(5) The degree of *correlation between the opposites test and related tests*, as measured by Aikins, Thorndike, and Hubbell, is shown in Table 53.

Notes.—Another relatively easy, though strictly controlled, association test is that of the backward-alphabet. This may be conducted in the same manner as the tests just described, by asking *S* to name, or to write as rapidly as possible, the letters that precede *f, k, s, p, w, l, e, r, d, o, v, j,n, t,* and *h.* For comparison, and to obtain a rough notion of *S*'s familiarity with the sequence of the alphabet in general, this test might be supplemented by another in which *S* was required to name or to write as rapidly as possible the letters that *follow* another series of 15 letters. For a written, group test, Aikins, Thorndike, and Hubbell allowed 15 sec. for the backward-alphabet test.

The same investigators combined the performances of *S*'s in the

TABLE 53

Correlation[1] of Opposites Test with Other Tests (Aikins, Thorndike and Hubbell)

FIRST MEMBER	SECOND MEMBER	CORRELATION IN PER CENT
List C. Number done.........	List B. Number done..........	50
List C. Number done	Alphabet. Number done.......	19-22
List B. Number done.........	Alphabet. Number done......	40-50
List B. Percentage of errors...	List C. Percentage of errors...	20
List B. Percentage of errors..	Alphabet. Percentage of errors	40
List C. Percentage of errors...	Alphabet. Percentage of errors	10
List B. Net efficiency.........	List C. Net efficiency.........	40
List B. Net efficiency.........	Alphabet. Net efficiency.......	60
List C. Net efficiency.........	Alphabet. Net efficiency.......	34

easy-opposites, hard-opposites, and backward-alphabet tests by ascertaining for each *S*: (1) the total number done (associations made), (2) the total number of errors, (3) the ratio of (2) to (1), (4) the total number of correct associations, and (5) the total net efficiency (by adding the efficiency scores in the hard-opposites and alphabet tests to one-half the net efficiency in the easy-opposites test).

REFERENCES

(1) H. A. Aikins, E. L. Thorndike, and Elizabeth Hubbell, Correlation among perceptive and associative processes, in P. R., 9: 1902, 374-382.

(2) Naomi Norsworthy, The psychology of mentally deficient children, New York, Nov.. 1906. Pp. 111.

(3) E. L.Thorndike, Educational psychology. New York, 1903. Pp. 173.

(4) E. L.Thorndike, An introduction to the theory of mental and social measurements, New York, 1904. Pp. 210.

(5) E. L.Thorndike, Measurements of twins, in Arch. of Phil., Psy., etc. No. 1, Sept., 1905 (Columbia Contr. to Phil. and Psych., 613 No. 3). Pp. 64.

TEST 35

Controlled association: Computation.—The solution of simple arithmetical problems in addition, subtraction, multiplication, and

[1] The correlations represented in this Table were figured by a special method, illustrated in Aikins (1: p. 381)

division may be considered as essentially dependent upon the accuracy and rapidity with which the appropriate associative processes are executed. Computation is, therefore, a test of controlled association in which the restriction of the associative sequence is complete, in which only a single outcome is correct. But numerous subsidiary activities are, of course, involved. Thus, the solution of arithmetical problems with the aid of paper and pencil demands, besides associative activity, both visual perception and motor activity, while mental computation imposes an additional tax by necessitating the holding in mind of the problem itself and of the various steps in its solution.

Because of this implication of perception, movement, attention, retention, and perhaps other forms of mental activity, as well as simple associative activity, the computation test has been employed not merely for the special purpose of studying the nature and course of associative processes, but also for the more general purpose of investigating mental efficiency at large (*geistige Leistungsfähigkeit*). Oehrn (19), for example, who was one of the first to use computation as a mental test, sought to study individual differences in the nature of associative processes, Aikins, Thorndike, and Hubbell (1) and Krueger and Spearman (15) to study the correlation of specific mental functions, Thorndike (27) to determine the relative influence of heredity and environment upon mental efficiency, Reis (21) to compare the ability of normal, paralytic, and hebephreniac children, Jones (10) to investigate the effect of bodily posture, and Vogt (29) the effect of distraction, upon mental efficiency. But the commonest application of the computation test has been made in the formulation of the curve of mental efficiency, or the work-curve (*Arbeitskurve*), with special reference to the influence of practise, rest-pauses, exercise, and similar factors upon the mental efficiency of children during a school-day. This use of the test is illustrated in the work of Bellei(2), Burgerstein (3), Ebbinghaus, (6) Friedrich (8), Holmes (9), Keller (11), Kemsies (12), Laser (16), Marsh (17), Schulze (22), Teljatnik (24), Thorndike (25), and Winch (28), as well as by numerous investigations in Kraepelin's laboratory.[1]

[1] See for this work the various volumes of the *Psychologische Arbeiten*.

Addition, multiplication, and both in alternation, have been more popular forms of computation than subtraction or division. With all four forms, varied types of problems have been used. These variations in the arrangement of the test naturally affect its outcome. The most important types of test are illustrated herewith. Beside the types that are shown, Winch employed miscellaneous arithmetical problems, while Reis had his S's add mentally for 1 min. by 7's or by 12's.

Examples of Material Used in Computation Tests

A	B		
2	4		
6	1		
9	3		
7	5	D	G
4	2	95799	93
8	6	86967	68
9	4	32687	41
5	3	84799	25
1	4	95976	52
7	6	34797	
5	1	97864	
2	5	98945	
6	4	87824	
3	2	68792	
5	1	79867	
1	2	88896	
3	3	97745	
9	6	39799	
2	5	48970	
1	3	89043	
3	6	67354	
6	5	54628	
9	1	91176	
8	2	9025	
3			
5			
2			
7			
6			
2			

C 4 2 8 3 2 9 9 5 4 6 5 4 3 1 7
7 9 2 9 3 8 3 8 2 6 5 5 1 3 9
1 1 0 2 etc.

E 42
+79

F 492
+763

H 64293643194831457627
38682725423585791858

I 982
−469

J 64
−27

K 28
× 8

L 364
×6

M 47
×89

N 948
× 579

O 7986
× 4523

R 4)799

P 428423995479253314325
4

Q 254)4659234(

EXPLANATORY

Addition

A. Vertical series of 1-place numbers, arranged to avoid repetitions and pairs adding to 10. *S*'s drop back to units when each hundred is reached. Oehrn, Vogt, and others working under Kraepelin used columns of as many as 7,000 of such digits. Krueger and Spearman used 70, grouped by 10's as illustrated. Marsh used similar columns of 15 or 25 digits.

B. Vertical column of 24 1-place numbers, using 1 to 6 only. Used by Jones, who had *S* add aloud while he himself followed with a check list.

C. Horizontal series of 1-place pairs of digits. A modification of the Kraepelin 1-place series in order to make possible the examination of the accuracy of each addition. The right-hand figure of the sum is the only one recorded, as illustrated in the first four problems. Used by Schulze, and apparently also by Ebbinghaus and by Vogt.

D. Twenty 5-place numbers. Used by Thorndike.

E. Two 2-place numbers. Used by Teljatnik.

F. Two 3-place numbers. Used by Kemsies for mental addition.

G. Five 2-place numbers. Twenty such problems were given and 2 min. allowed for computation. Used by Thorndike and by Aikins, Thorndike and Hubbell.

H. Two 20-place numbers. Used by Burgerstein, Laser, Friedrich, and Holmes. The last-named investigator published elaborate rules for the construction of these problems in such a way as to avoid the extension of errors in 'carrying.' She used 4 blanks with 16 such problems on each blank.

Subtraction

I. Two 3-place numbers. Used by Kemsies for mental subtraction.

J. Two 2-place numbers, to be written on the blackboard (Teljatnik).

Multiplication

K. Two-place multiplicand, 1-place multiplier. Used by Kemsies for mental computation, and by Ebbinghaus for written group tests.

L. Three-place multiplicand, 1-place multiplier. Used by Kemsies.

M. Two-place multiplicand, 2-place multiplier. Used by Keller, and by Marsh with the digits 1, 2, 5, and 9 excluded.

N. Three-place multiplicand, 3-place multiplier. Used by Keller for written, and by Thorndike for mental computation.

O. Four-place multiplicand and multiplier. Used by Thorndike both for written, and for mental computation. The multiplicand was usually a combination of 6, 7, 8, and 9; the multiplier of 2, 3, 4, and 5.

P. Twenty-place multiplicand, 1-place multiplier. Used by Burgerstein, Laser, and Friedrich, with the restriction of the multiplier, in most tests, to 2, 3, 4, 5, or 6.

Division

Q. Three-place divisor, 7-place dividend. Four blanks of 10 problems each were used by Bellei for an hour's work.

R. One-place divisor, 3-place dividend. Used by Kemsies for mental computation.

There are certain advantages and certain disadvantages in each of these forms of material. In general, *E* must select that form of test that best suits the conditions under which he works.

MATERIALS.—Watch. Printed blanks, containing problems in addition or multiplication.

Five forms have been prepared for this test: others may be prepared by *E* as desired.

A. Addition test: several thousand digits in vertical columns with a line separating each 10 digits, after Model A. This form may be used with children or with adults, and either for short series or for continuous adding, after the Kraepelin method or after the plan of Krueger and Spearman.

B. Addition test with 36 problems, patterned after Model G, but containing 10, in place of 5 numbers each.

C. Addition test, patterned after Model C (Schulze's method), and specially recommended for younger *S*'s.

D. Addition test, patterned after Holmes, Model H, and virtually identical with the material used by Burgerstein, Laser, and Friedrich.

E. Multiplication test, after Model P, as used by Burgerstein, Laser, and Friedrich.

METHOD.—For individual tests, Form B may be used with oral solution. Supply *S* with the printed test-sheet. Instruct him to call out the sums as rapidly as possible, but to correct any errors consciously made. Take his time for the solution of all the problems, and, if desired, check off upon another sheet the point he has reached at every 10 or every 30 sec.[1] This test has the advantage of freeing *S* from the labor of recording his results—a task which tends to introduce errors into the work of young, and to check the time of older *S*'s.

[1] A refinement employed by Kraepelin enables *E* to record the time of every addition. *S* writes the sums, using a pencil with electrical contacts (see Schulze, 23: p.250, or P. A., 2: 1899, 400), and a graphic record with time-control is arranged.

Form A may be added orally by sections of 10 digits each. The other forms are less well adapted for oral solution, and, as the work is somewhat more difficult, it is not likely that *S* can compute faster than he can record his results.

Reis' method may also be employed to advantage.

For group tests, work by the time-limit method, selecting such a time that the fastest S in the group to be tested can no more than complete the task.

If the group test is to be used with school children to determine efficiency under various conditions or at various times of the day, *E* must arrange the experiment to exclude, or at least to measure the influence of, all possible disturbing factors. The most serious of these are practise, excitement, ennui, and carelessness.

A common method for cancelling out practise is to divide *S*'s into two equivalent groups on the basis of a preliminary test, and to administer one set of problems early to the first, and late to the second group (if, for instance, fatigue is to be investigated), the other set late to the first, and early to the second group.

In studying the work curve, some *E*'s have used computation both as the test and as the work to induce fatigue, practise, etc.; others have used computation as a test of efficiency, but have allowed *S* to follow in the main the regular work of the school session. In the first procedure, computation (usually addition) is pursued more or less continuously for an hour, or even for several hours; in the second procedure, the computation itself occupies but a short time, relatively, say from 1 to 10 min., and is repeated at intervals of an hour or more, while *S* meantime takes up his regular tasks, indulges in physical activity, or rests, as *E* may direct.

In illustration, Vogt, Oehrn, and other disciples of Kraepelin, have kept their *S*'s adding continuously for several hours; Holmes used 4 periods of adding of 9 min. each, with 4-min. rest-pauses, Burgerstein 4 periods of 10 min. each, with 5-min. pauses. Typical illustrations of the second procedure are supplied by the investigations of Laser and of Ebbinghaus, who introduced 10-min. computation tests at the beginning of the school day and once an hour thereafter. Ebbinghaus is inclined, however, to recommend 5-min. tests as being equally serviceable for the determination of efficiency and less likely to develop ennui and carelessness. Offner (20, pp. 30-31) favors short tests for similar reasons and also for the partial avoidance of the practise-error.

Treatment of Data.—Computation tests yield two indexes of efficiency—speed or quantity of work and accuracy or quality of work. As a rule, no attempt has been made to unite these two indexes into a single index of net efficiency, although this might be done as explained in the Cancellation Test. Some *E*'s, *e.g.*, Oehrn, have considered speed of work only; others, *e.g.*, Teljatnik, quality of work only.

Quantity of work is indicated by elapsed time when using the individual method, and by the number of problems solved (sometimes by the number of figures written in the results) in the time-limit or group method.

Quality of work is generally regarded as directly proportional to the percentage of correct solutions. Inaccuracy is most often taken in terms of the number of errors committed, less often in terms of the number of errors plus the number of corrections made by *S*. The simplest, but the least desirable way to compute errors is to score one error for every wrong figure in the result. In the case of certain problems, however, a single error in computation may affect more than one figure in the result.[1] For reliable results, these complex errors must be examined and the score adjusted to indicate exactly the number of real errors of computation.

Results.—(1) There are marked *individual differences* in the speed and accuracy of computation, even among *S*'s of the same age and same school class. Schulze's best pupil added more than 5 times as fast as the slowest pupil in the same class. Similar differences were discovered by Keller with school children and less striking differences by Oehrn with adults.

(2) Where *sex differences* have been noted, the general evidence is in favor of a slight superiority of girls over boys, *e.g.*, in the results of Miss Holmes. That the difference is but slight may be inferred from the conclusions of Fox and Thorndike (7) that the girls in the high school studied did about 5 per cent better on the whole than the boys, but this conclusion is offered with the reservation that a better grade of girls was probably selected by the school. Similarly, in combined ability with the opposites, alphabet, addition, multiplication, and word tests, as reported by

[1] The problems in Form D (Example H) are intentionally arranged to reduce this error.

Thorndike (26, p. 117), 48 per cent of the boys reached the degree of ability reached by 50 per cent of the girls.

(3) *Posture* was found by Jones (10) to affect the speed of adding; both children and college students could add somewhat faster (approximately 3 to 8 per cent) with the body in a horizontal than with the body in a vertical position.

(4) The *effect of distraction* by concomitant activities, *e.g.*, the reciting of a poem, was found by Vogt (29) to reduce very materially (58.5 per cent) the number of additions made by the continuous (Kraepelin) method, but to have relatively little effect upon the simpler process of adding pairs of digits.

(5) *Practise* increases the speed of computation, even when the test demands the exercise of associative connections in which *S* has already had extended school training. The interpretation of practise-effects in the analysis of the work curve is treated more fully below.

(6) The results of investigations of the *work curve* may be most conveniently discussed under several distinct rubrics, since the conditions under which the tests have been applied appreciably affect their outcome. In particular, we may profitably observe the distinction already drawn between investigations in which computation has served both as test and as work, and investigations in which computation has served as test, but other forms of mental activity as work proper. The relation of individual differences to the general results of these investigations needs special consideration, as does also the effect of the insertion of rest-pauses.

(*a*) *General analysis of the work curve.* We have had occasion already[1] to discuss the principal factors that condition the work curve. Extended accounts of these factors will be found in Meumann (18, II., 8 ff. and elsewhere), Schulze (23, 251 ff.), and Kraepelin (13, 14). The last-named investigator has been most active in this field, and to him we owe most of our information with regard to the part played by practise, fatigue, recuperation, adaptation (*Gewöhnung*), momentum or 'warming-up' (*Anregung*), and the various kinds of spurts (*Antriebe*). In the experimental study of school children, still another factor, loss of interest, or ennui, and

[1] See especially the Tapping and the Cancellation Tests, Nos. 10 and 26.

resultant carelessness, enters as a serious source of disturbance. It is clear that to determine the rise and development of any one of these factors, of fatigue for instance, the other factors must be eliminated or evaluated.

(*b*) *Individual differences in the work curve.* We have noted the presence of clear-cut individual differences in the speed and accuracy of computation; there are also individual differences in the course of the performance. Thus, both Kemsies and Keller conclude that mass results should be subjected to scrutiny to detect individual curves of performance, if reliable information is to be secured concerning fatigue and overpressure in the schools.

Despite the great variety of individual work curves, they may be classified into a small number of characteristic patterns. Kemsies, for example, distinguishes between persistent workers who fatigue slowly and profit much by practise and feeble workers who fatigue quickly and do not profit much by practise. Meumann's own investigations lead him (18: vol. 2, pp. 10–11) to posit three types of workers (quantitatively regarded): the first type attains maximal efficiency at the start and thence decreases with many fluctuations: the second attains maximal efficiency only after an interval (of a length depending upon the kind of work); the third attains maximal efficiency only after a long period, perhaps several hours, of work. The first type, then, is characterized by rapid adaptation and rapid fatigue, the second by slower adaptation and slower fatigue, the third by very slow adaptation and very great resistance to fatigue. The third type is probably more common in adult males, the first in women and children.

(*c*) *The work curve for continued computation.* 1. Work without interruption. Oehrn found that when adults added continuously for 2 hours or more, maximal speed was attained on the average at about 28 min. from the start.

Schulze finds, however, that with school girls aged 12.5 yrs., signs of fatigue appear even in the first 5 min. The total number of additions made per minute by 37 girls was 1850, 1871, 1863, 1785, and 1772 for the 1st to the 5th minute, respectively.

Schulze's results with the same pupils for longer periods (50 min. without pause) show a progressive decrease both of quantity and of quality of work (Table 54). These figures, which are selected

from the 6th of a series of experiments, are based upon the very easy process of adding two 1-place digits. Practise has, therefore, relatively little effect, but fatigue diminishes efficiency.

TABLE 54

Efficiency in Addition: Five 10-Minute Periods (Schulze)

PERIOD OF TEN MINUTES	TOTAL NUMBER OF ADDITIONS	PERCENTAGE OF DECREASE OF QUANTITY OVER THE PREVIOUS PERIOD	PERCENTAGE OF DECREASE OF QUALITY OVER THE PREVIOUS PERIOD
I..............	17,740	—	—
II..............	16,726	5.7	.09
III..............	15,855	5.2	.03
IV..............	15,485	2.3	.17
V..............	15,134	2.3	.01

2. Work with interruptions. When repeated computation tests are made within an hour, the usual result is a progressive increase in the quantity, but a progressive decrease in the quality of the work. Burgerstein's figures (Table 55) furnish a typical example of the results for four 10-min. periods with 5 min. rest-intervals between periods.

TABLE 55

Efficiency in Addition and Multiplication within an Hour (Burgerstein)

PERIOD	NUMBER OF FIGURES IN RESULTS	NUMBER OF ERRORS	PERCENTAGE OF ERROR
I..............	28,267	851	3.01
II..............	32,477	1292	3.98
III	35,443	2011	5.67
IV..............	39,450	2360	5.98

Miss Holmes' results are similar, though, on account of computing errors of a different plan ('serial' errors counting but as one error), her percentage of error averaged but 1.3 as against Burgerstein's 3.[1]

[1] Miss Holmes' analysis of the errors showed that their increase during the hour was due primarily to increased inaccuracy in associative processes, rather than to increased frequency of 'slips of the pen.' In general, errors of transcription were about one-third as numerous as errors of association.

The general interpretation of results such as Burgerstein obtained has been that practise increases the speed of work, while fatigue increases its inaccuracy. Schulze's work with problems which were assumed to minimize the effect of practise may be regarded as corroborative of this view. On the other hand, Ebbinghaus (6, pp. 406-7) denies that practise could produce such marked increase of speed, and therefore ascribes both the increase in speed and the increase in inaccuracy primarily to progressively augmenting haste and carelessness.

(*d*) *The work curve with short and scattered computation tests.* Typical instances of the use of computation as a test for the fatigue-effects of the regular school program are afforded by the experiments of Friedrich, of Laser, and of Ebbinghaus. This method has been adopted in part to avoid the entrance of ennui and carelessness, as just mentioned.

Laser's tests at hourly intervals of 226 pupils (aged 9-13 yrs.) in a Königsberg Bürgerschule are summarized in Table 56. Inspection shows that, save for the 5th period, the outcome is the same as that of the tests for an hour's time by Burgerstein, viz: a progressive increase, both in speed and in inaccuracy of computation.

TABLE 56

Efficiency in Computation within a School Session (Laser)

TEST AFTER SCHOOL PERIOD	TOTAL NUMBER OF FIGURES ADDED	TOTAL NUMBER OF ERRORS	PERCENTAGE OF ERROR
I..............	34,900	1147	3.28
II..............	40,661	1460	3.59
III	43,124	1713	3.79
IV..............	43,999	1796	4.08
V..............	45,890	1668	3.63

Ebbinghaus, who sought to determine the desirability or undesirability of a 5-hour continuous school session in a Gymnasium and higher girl's school at Breslau, obtained results identical with those of Laser so far as the qualitative aspects are concerned, but differing somewhat as regards the quantitative aspects, more particularly in that speed of computation reached a maximum at the close of the 2d school period, to remain thereafter almost constant or to fall off slightly toward the close of the session

(*e*) *The effect of pauses upon the work curve.*[1] Periods of rest exercise a generally favorable effect upon efficiency in computation, especially upon the quality of work.

That a pause of 5 min. will favor the quantity of work done within a 50-min. period is shown by the data of Table 57, which are derived by Burgerstein from Schulze.

TABLE 57

Additions per Pupil, with and without a Rest-Pause (Burgerstein-Schulze)

	FIRST 25 MINUTES	REST-PAUSE	SECOND 25 MINUTES
First test........	1067	5 min.	1088
Second test......	1146	None	1042

Similar conclusions are reached by Kraepelin from his tests of adults (13, pp. 16-17).

The extended tests made by Friedrich upon 51 ten-year old Volksschule pupils included both exercises in dictation and in computation (20-min. tests with addition and multiplication). His results for computation tests are summarized in Table 58. Consultation of the original records will show a striking agreement between the data for dictation and those for computation.

(*f*) *Efficiency at different periods of the day.* Friedrich's results lead him to advise lighter work in the afternoon session. Bellei (2) found that boys and girls aged 12 solved problems in division more slowly and less accurately in the afternoon than in the morning. Marsh (17) tested but a few individuals, so that it is probably unsafe to make inductions from his data, which seemed to indicate a greater efficiency in adding at noon than later in the day, and in multiplication at between 1.30 and 3 p.m. than at 6 or at 10.30 p.m. With adults there are certainly marked individual differences in the diurnal ebb and flow of efficiency. Frequently quoted are the experiments conducted by Thorndike (25) upon

[1] For a full discussion of the experimental examination of rest-pauses in their relation to efficiency, consult Burgerstein and Netolitzky (4, pp. 554-565).

TABLE 58

Effect of Pauses upon Computation (Friedrich)

	TEST	TIME	PAUSE	ERRORS PER THOUSAND
FORENOON	I.......	Before 1st hour	—	11.19
	II.......	After 1st hour	None	16.38
	III.......	After 2d hour	8 min. after 1st hour	19.54
	III*a*......	After 2d hour	None	20.26
	IV.......	After 3d hour	Two of 15 min. each after 1st and 2d hours	19.36
	IV*a*	After 3d hour	One of 15 min. after 2d hour	22.28
	IV*b*......	After 3d hour	None	22.76
AFTERNOON	V.......	Before 1st hour	—	18.88
	VI.......	After 1st hour	None	20.61
	VII.......	After 2d hour	15 min. after 1st hour	21.04
	VII*a*	After 2d hour	None	24.06

adults and children at different times of the day, both when the *S*'s were fresh and well rested and when they reported feelings of weariness. The net conclusions drawn by this investigator are that "incompetence, mental fatigue, does not come in regular proportion to the work done," that feelings of fatigue are not measures of mental inability, that disinclination to work does not signify inability to work. Thus, in schoolroom tests at Cleveland, Ohio, and Scranton, Pa., pupils tested late (between 10 and 40 min. before school closed) did 99.3 per cent as much work and made 103.9 per cent as many mistakes as did those tested early (10 to 40 min. after school opened). It may be disputed, however, whether this demonstration that pupils *can* work nearly as well at the end of school session as at its beginning is equivalent, as some writers have thought, to a demonstration that they should be expected to work as well at the later periods.

A special study of *fatigue in evening schools* by Winch (28) leads him to the conclusion "that evening work is comparatively unprofitable and that a short time in class in the evening is sufficient, *plus* the labors of the day, to induce a low condition of mental energy." Illustrative results are given in Table 59.

TABLE 59

Effects of Fatigue on Arithmetical Work in Evening Schools (Winch)

TEST	GROUP-A OBSERVERS		GROUP-B OBSERVERS	
	Time	Mean Score	Time	Mean Score
I..............	8.30	302	8.30	320
II..............	8.00	363	9.00	286

(*g*) *Efficiency on different days of the week.* The work of Kemsies seems to indicate that efficiency in school children is higher in the first than in the later days of the week.

(*h*) *Efficiency at different periods of the year.* Bellei concludes that a poorer quality of work is done by school children in June than in January, because of the accumulated effect of a long period of school work.

(7) *Correlations.* Aikins, Thorndike, and Hubbell compared efficiency in adding with efficiency in the other 'association' tests (misspelled words, cancellation of two letters, and opposites), and (by a special method of estimating the index) found the quality of work in adding and quantity of work in associating correlated to a degree of 50 per cent in 8th-grade, and 20 per cent in 5th-grade pupils, and net efficiency in adding and net efficiency in associating correlated to a degree of 48 per cent. On the other hand, the percentage of error in adding and in the other association tests exhibited no correlation or one of but slight degree.

Thorndike's study of mental resemblances in twins (27) showed a much higher correlation of ability in computation between twins than between siblings; thus, twins aged 9–11 years revealed a correlation of 0.90 in adding, and 0.91 in multiplication, and twins aged 12–14 years a correlation of 0.54 in adding and 0.69 in multiplication: taken collectively, the index of correlation amounted to 0.75 for the adding, and 0.84 for the multiplication test.

Fox and Thorndike (7) found that ability to add correlated to a fairly high degree, 0.75, with ability to multiply, but only to a small degree, 0.20 to 0.44, with ability to solve fractions or to perform other arithmetical problems. They conclude that "ability in

arithmetic is thus but an abstract name for a number of partially independent abilities." Similarly, Burris (5) found that school grades in algebra and in geometry, as recorded in 19 representative high schools, showed, for nearly 1000 pupils, a correlation of but 0.45.

Krueger and Spearman (15, pp. 77–78) found a correlation between adding and pitch discrimination of 0.83, and between adding and the Ebbinghaus completion test of 0.93.

(8) *In mental defectives,* ability to compute has been investigated by Reis, who found the average performance less and the variability greater in the work of paralytics and hebephreniacs than in normal individuals.

(9) *Relation of quality and quantity of work.* In general, for a given *S*, speed and accuracy of computation are related inversely, but this generalization suffers exceptions. Kemsies found that the period of fastest computation did not coincide with the period of most accurate computation, a result which is illustrated in several of our tables. In comparing individual results, he found, similarly, that those who worked most rapidly tended to work most inaccurately.

(10) *Qualitative analysis.* Those who have used computation tests have not sought, as a rule, to examine the mental processes involved in them. Oehrn, however, calls attention to the fact that practise in adding (by the Kraepelin method) tends to induce quasi-automatic addition. This circumstance, taken in conjunction with the relatively small correlations between different forms of computation themselves, and between them and other abilities, lends countenance to Well's objection (30) to accepting the computation test, without further qualification, as a measure of general mental efficiency.

REFERENCES

(1) H. Aikins, E. L. Thorndike and Elizabeth Hubbell, Correlations among perceptive and associative processes, in P. R., 9: 1902, 374–382.

(2) Bellei, Ulteriore contributo allo studio della fatiga mentale nei fanciulli, in Rivista sperim. di freniatria, 30: May, 1904. See summary by Binet in A. P., 11: (1904) 1905, 369.

(3) L. Burgerstein, Die Arbeitskurve einer Schulstunde, in Z. S., 4: 1891, 543–564, 607–627. Also published separately in German, and in condensed form in English as: The working curve of an hour.

(4) L. Burgerstein and A. Netolitzky, Handbuch der Schulhygiene, 2d. ed., Jena, 1902. Pp. 997.

(5) W. P. Burris, The correlations of the abilities involved in secondary school work, in C. C., 11: 1903, 16–28.

(6) H. Ebbinghaus, Ueber eine neue Methode zur Prüfung geistiger Fähigkeiten in ihre Anwendung bei Schulkindern, in Z. P., 13: 1897, 401–457.

(7) W. S. Fox and E. L. Thorndike, The relationships between the different abilities involved in the study of arithmetic. Sex differences in arithmetical ability, in C. C., 11: Feb., 1903, 32–40.

(8) J. Friedrich, Untersuchungen über die Einflusse der Arbeitsdauer und der Arbeitspausen auf die geistige Leistungsfähigkeit der Schulkinder, in Z. P., 13: 1897, 1–53.

(9) Marion E. Holmes, The fatigue of a school hour, in Pd. S., 3: 1895, 213–324.

(10) E. E. Jones, The influence of bodily posture on mental activities, N. Y., 1907. Pp. 60. (Reprinted from Archives of psychology, No. 6.)

(11) R. Keller, Ueber den 40-Minutenunterrichtsbetrieb des Gymnasiums u. der Industrieschule in Winterthur, in I. M. S. H., 2: 1906, 298–330, especially 307–318.

(12) F. Kemsies, Arbeitshygiene der Schule auf Grund von Ermüdungsmessungen, in S. Z., 2: 1898. Pp. 64.

(13) E. Kraepelin, Ueber geistige Arbeit, 2d. ed., Jena, 1897. Pp. 29.

(14) E. Kraepelin, Die Arbeitscurve, in Ph.S., 19: 1902, 459–507.

(15 F. Krueger and C. Spearman, Die Korrelation zwischen verschiedenen geistigen Leistungsfähigkeiten, in Z. P., 44: 1907 50-114.

(16) H. Laser, Ueber geistige Ermüdung beim Schulunterrichte, in Z. S., 7: 1894, 2–22.

(17) H. D. Marsh, The diurnal course of efficiency (Columbia Univ. thesis), N. Y., 1906. Pp. 99.

(18) E. Meumann, Vorlesungen zur Einführung in die exp. Pädagogik, 2 vols., Leipzig, 1907. Pp. 555 and 467.

(19) A. Oehrn, Exp. Studien zur Individualpsychologie, in P. A., 1: 1896, 92–151.

(20) M. Offner, Die geistige Ermüdung, Berlin, 1910. Pp. 88.

(21) J. Reis, Ueber einfache psychologische Versuche an Gesunden u. Geisteskranken, in P. A., 2: 1899, 587-694.

(22) R. Schulze, 500,000 Rechenaufgaben. Eine exp. Untersuchung, in Praktische Schulmann, 44: 1895, 340.

(23) R. Schulze, Aus der Werkstatt der exp. Psychologie u. Pädagogik, Leipzig, 1909. Pp. 292.

(24) Teljatnik, article in Russian. See for details, Burgerstein u. Netolitzky, 4: especially pp. 462–5.

(25) E. L. Thorndike, Mental fatigue, in P. R., 7: 1900, 466–482, 547–579.

(26) E. L. Thorndike, Educational psychology, N. Y., 1903. Pp. 177.

(27) E. L. Thorndike, Measurements of twins, in Archives of Phil., Psych. etc., No. 1, Sept., 1905. Pp. 64.

(28) W. H. Winch, Some measurements of mental fatigue in adolescent pupils in evening schools, in J. E. P., 1: 1901, 13–23, 83–100.

(29) R. Vogt, Ueber Ablenkbarkeit u. Gewöhnungsfähigkeit, in P. A., 3: 1901, 62–201, especially 80–118, 131–5.

(30) F. L. Wells, Technical aspects of experimental psychopathology, in Amer. J. of Insanity, 64: January, 1908, 477–512.

TEST 36

Learning: Habit-formation in mirror-drawing.—The preceding tests of association deal with *S*'s facility in producing unrestricted series, or in reproducing restricted series that have already been learned. The present test compels *S* to form a new series of associations that are opposed to associations stereotyped by several years of daily experience.

More particularly, in tracing an ordinary drawing, the movements of the hand are guided by the visual perception of the drawing, plus kinesthetic sensations set up by the movement of the pencil. If the drawing is seen in a mirror, the natural relations are reversed in certain respects, so that a new series of associative connections must be established between eye and hand. The rapidity and ease with which these new connections are established is taken as an index of learning-capacity.

Learning is often said to take place either by practise (trial and error), by imitation, or by some form of ideational control (instruction, reasoning, etc.). In the mirror-drawing test, the conditions preclude the use of imitation, and there is but relatively little opportunity to employ ideational control; whatever improvement appears is due primarily to a process of trial and error.

Mirror-drawing has been suggested by Judd (5, p. 99) to determine the effect of practise in the acquisition of motor habits, by Starch (8) as a demonstration-experiment to illustrate the trial and error method of learning and the cross-transfer of practise, and by Burt (3) to measure adaptability[1] and its correlation with general intelligence.

[1] For a discussion of prolonged adaptation to mirror-vision, see Stratton (9).

APPARATUS.—A 6-pointed star, printed in red ink.[1] Mirror. Cardboard about 17 x 24 cm. [For the subsidiary tests, a 20 deg. prism.]

PRELIMINARIES.—Pin the star out flat upon a table, directly in front of *S*, with the small cross-line that indicates the starting-point at the back, and the card square with the edge of the table. (This brings the star slightly 'out of true,' as is intended.) Set up the mirror, inclined slightly (about 5 deg.) from the vertical, just beyond the card. Arrange the screen (see Fig. 52) so that it will cut off *S*'s direct view of the star, but will allow him to see it clearly in the mirror, and will not interfere with his hand in drawing.

METHOD.—Place the point of a lead pencil upon the cross-line of the star, and assist *S* to grasp the pencil (permitting him to look only in the mirror). Instruct *S:* "Trace the outline of the star, starting in this direction [indicating, *by pointing*, the tip of the star at the right of the cross-line]. Work as rapidly as you can, but try to keep on the line. Don't stop to figure out what you ought to do, but keep your pencil going in some direction, and keep its point on the paper all the time." Start the watch, and record the time for the entire drawing.

E may also note the time for each sixth of the pattern. But it is, perhaps, more desirable to supplement the total time by a record of the total number of corrective movements made by *S*. Since these movements are often rapid, and of short extent, it is necessary to use a mechanical, or other form of counter to obtain the record. Press the counter every time *S* moves toward the line.[2]

[1] This star resembles the pattern devised by W. F. Dearborn, and described by Starch, but it has been modified by being made slightly larger, printed in red ink, and tipped somewhat away from the vertical position, in order to avoid the too-easily drawn vertical lines.

The author is experimenting with still another modification, viz: the use of double concentric stars, in the hope that this pattern may restrict the extent of permissible excursion from the printed outline. With the single star, some *S*'s are inclined to slight their work, and so to reduce their time, while others try to keep exactly upon the line, with consequent extension of their time-records. Thus, one *S*, on his first trial, followed the tracing very poorly but completed his work in 46 sec.: on being asked to keep nearer the line, he consumed, on his second trial, 195 sec.

[2] Every 'error,' or movement away from the line must, of course, be compensated for by a return movement. The idea is to register the number of these errors, or corrective movements. Changes of direction necessitated by the pattern itself are, obviously, to be neglected.

For a standard test, make 5 trials with the right hand, using a fresh star for each trial.

VARIATIONS OF METHOD.—Make the first trial with the left hand: follow with a series of from 5 to 50 trials with the right hand: return to the left hand for a final test. Note how much practise effect has been 'transferred' from the one hand to the other. Plot a graph to show the effect of practise, both upon the time and upon the corrective movements.

FIG. 52. THE MIRROR-DRAWING TEST.

TREATMENT OF DATA.—In the standard form of test, E has available 5 records, in each of which he knows the total time and the number of corrective movements. S may then be compared with other S's in respect to his performance in the first trial, or in all five trials, or in respect to the relative gain in the 5th as compared with the 1st trial, and, in each case, either total time or number of corrective movements may be made the basis of comparison.[1]

[1] Other methods of treatment are suggested by Burt, who, however, concluded that the second of those we have mentioned (average or sum of speeds in all the trials) was most practical.

Results.—(1) *Individual differences* in this test are manifest: the time for tracing the star at the first attempt varies, for adults at least, from about 50 secs. to over 8 min.

(2) *Sex.* The author's tests, as well as those of Burt (by a somewhat different method), indicate faster speed for women and girls than for men and boys. It is not only possible, but probable, that this difference is due in large measure to greater familiarity of women with the use of the mirror, though Burt believes that it may be due in part, also, to an innate sex difference. Illustrative data are given in Table 60: other tests with adults have revealed the following sex differences:

Eleven men, average time 220 sec., average corrective movements 75.

Eight women, average time 130 sec., average corrective movements 34.

(3) (*a*) The *effect of practise* for a short series is shown in Table 60, and for a series of 100 trials, one trial daily, in Fig. 53.

TABLE 60

Effect of Practise on Speed in Mirror-Drawing. College Students (Whipple)

	Number	1st Left	1st Right	2d Right	3d Right	4th Right	5th Right	2d Left
Men.......	11	169	127	108	96	80	67	88
Women...	23	149	127	87	76	67	67	74

(*b*) In Starch's long period of practise, it will be noted that the improvement in speed and in accuracy (freedom from corrective movements) is not parallel, save for the first 7 or 8 days. Both speed and accuracy seem to have reached their limit of improvement at the 90th day.

(*c*) *Cross-education.* A considerable amount of practise gained with the one hand is transferred to the other (unpractised) hand. Thus, Starch's 100-day practise with the right hand effected an improvement in it of 92 per cent in accuracy and of 84 per cent in speed. A single left-hand record, made at the expiration of this period, showed, in comparison with a single left-hand record made before practise began, an improvement of 85 per cent in speed and of 81 per cent in accuracy. There is, however, nothing surpris-

ing in this so-called 'cross-education,' since the tracing of the star in the mirror depends primarily upon coördinations established in the central nervous system: in other words, the transfer is only an outwardly apparent transfer: in reality, the same factors are at work in the control of either hand.

(*d*) *Persistence of practise.* The effect of even a short period of practise in mirror-drawing is very persistent. Thus, Burt administered 6 tests in succession, during which the average speed fell from 103 to 39.5 sec. Twelve weeks later, two tests were given

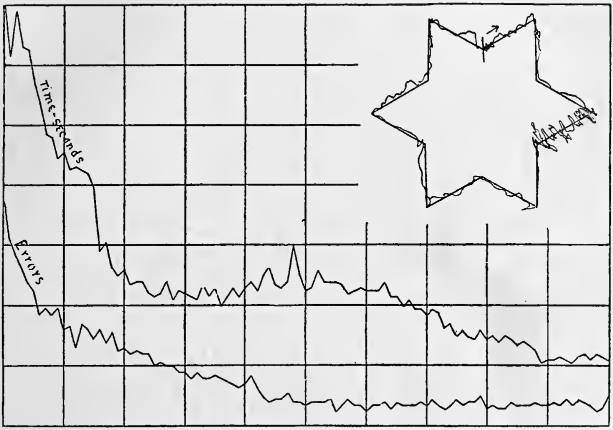


FIG. 53. THE EFFECT OF PRACTISE UPON MIRROR-DRAWING. (From Starch)

in succession; the average speed developed was 34.5 sec. in the first, and 27.4 sec. in the second: in other words, the 7th test surpassed the 6th, made 12 weeks previously.

The extent to which this persistence of practise-effect was shared by Burt's *S*'s is indicated by the correlation 0.52 between their standing before, and their standing after the 12-week interval.

(4) *Relation of speed and accuracy.* We have already noted that speed and accuracy (freedom from corrective movements) do not necessarily follow parallel courses as practise continues. Yet, in

general, these two indexes of efficiency are closely correlated. The author's data for 19 cases indicate, for example, a correlation of 0.86, P.E., 0.04, between total time and number of corrective movements.

(5) *Correlation with intelligence.* The author could discern no constant differences between the work of 5 dull and 5 bright boys. Burt, however, found a correlation between speed and estimated intelligence of 0.67, P.E. 0.07, for elementary school boys, and of 0.54, P.E. 0.14, for preparatory school boys.

(6) *Qualitative aspects.* Efficiency in mirror-drawing may result from the actual formation of new visual-motor coördinations (indeed, some *S*'s after executing a number of drawings, find that, for a short time immediately thereafter, these new coördinations interfere with normal drawing or writing); but efficiency may also result from the voluntary inhibition of visual control in favor of kinesthetic control, *i.e.*, by thinking the drawing of a star in motor terms, as if working with the eyes shut. Or, the hand-movements may be started in this manner and then carried out by visual control from the mirror. Finally, adults occasionally control the drawing ideationally, *i. e.*, by applying inferred properties of reflection by mirrors.

It is evident that the existence of these qualitative differences may invalidate the test, in so far as the quantitative data for different *S*'s may 'measure' radically different mental processes.

Very slow *S*'s get 'caught' at certain difficult points of the drawing, where they make a long series of futile attempts to start in the right direction. Here the normal visual-motor control is too persistent to be readily broken or ignored.

NOTES.—A further study of the associative connections involved in mirror-drawing may be made by the use of dot-tapping through a prism or of the various forms of mirror-writing.[1]

[1] On mirror-writing, consult Abt, Allen, Downey, Laprade, Lochte, Weber, and Wegener. The most elaborate statistical study is that of Lochte, who examined 2804 pupils in Berlin, and found, for children aged 6-7 years, 13.2 per cent of spontaneous left-hand mirror-writing in boys and 25.4 per cent in girls, but for children aged 13-14 years, only 0.7 per cent in boys and 35 per cent in girls. The tendency toward this type of writing appears, therefore, to decrease with age, and to be more evident in girls than in boys.

The most elaborate qualitative analysis of the various 'controls' used in writing is that of Miss Downey.

For the first test, let S shut his left eye, and strike repeatedly with his right forefinger, at a mark on the wall or table-top, making about one stroke per second, after the manner prescribed in the test of precision of aiming (No. 11). After this rhythmic movement has become well established, and *without interrupting it in the least,* place suddenly before his eye a 20 deg. prism, with the base toward his nose. The mark is thereby apparently displaced some 10 cm. to the left. Count the number of strokes that S makes before he hits the mark again (with the prism kept before the eye). Similarly, count the number of strokes necessary to hit the mark again when the prism is removed.

For the second test, try any or all of the following:

(1) Close the eyes and write with both hands simultaneously. Cases will then appear, particularly in young children, of spontaneous mirror-writing (writing which reads correctly when held before a mirror) with the left hand. If this appears, see if S can write normally with the left hand when his eyes are closed.

(2) Show S a sample of mirror-writing. Explain its nature. Ask him to write in a similar manner, first with his left, then with his right hand.

(3) Write with both hands simultaneously, but with the left intentionally in mirror-writing.

(4) Read normal writing when seen only as reflected in a mirror.

(5) Write normally while watching the writing in the mirror, *i.e.*, with hand and paper hidden from direct observation, as in the star test.

REFERENCES

(1) G. Abt, L'écriture en miroir, in A. P., 8: 1901 (1902), 221–225.

(2) F. J. Allen, Mirror-writing, in Brain, 19: 1896, 385–7.

(3) C. Burt, Experimental tests of general intelligence, in B. J. P., 3: December, 1909, 94–177, especially 145–9.

(4) June E. Downey, Control processes in modified hand-writing: an experimental study, in P. R. M. S., 9: April, 1908, No. 37. Pp. 158.

(5) C. H. Judd, Laboratory manual of psychology, New York, 1907.

(6) A. Laprade, Contributions a l'étude de l'écriture en miroir. 1902.

(7) Lochte, Beitrag zur Kenntnis des Vorkommens u. der Bedeutung der Spiegelschrift, in Arch. f. Psychiatrie u. Nervenkrankheiten, 28: 1896, 379–410.

(8) D. Starch. A demonstration of the trial and error method of learning, in P. B., 7: January, 1910, 20–23.

(9) G. M. Stratton, The spatial harmony of touch and sight, in Mind, n.s. 7: 1899, 492–505.

(10) H. Weber, Spiegelschrift u. Lenkschrift, in Zeits. f. klin. Med., 27.

(11) H. Wegener, Die Spiegelschrift, in Z. P. P., 1: 1899, 254–269.

TEST 37

Learning: Substitution.—This test is one of many that may be devised to measure the rapidity with which associations are formed by repetition. Here the associations consist in connections between a series of symbols and the nine digits, and are established gradually as the test proceeds. In theory, an *S* whose nervous system is plastic and retentive, who, in other words, is a quick learner, will make most rapid progress. The learning differs in type from that of the proceeding test, since here the conditions demand that auditory, visual, or auditory-visual connections be formed by sheer repetition, whereas there, visual-motor connections were formed by trial and error.

The principle embodied in the test admits, as has just been said, of numerous variations in detail of application. The two forms here presented are themselves modifications of a test devised by W. F. Dearborn, and used for some time at the University of Wisconsin.

A. STANDARD FORM FOR INDIVIDUAL PROCEDURE

Materials.—Printed cardboard form, 3 x 7 in., containing 9 circles. Each circle contains a digit (from 1 to 9) and a symbol (square, asterisk, etc.). Sheet of 10-ply cardboard, 7 x 14 in. Four strips of cardboard, each 1 in. wide and 14 in. long. Test-strips of heavy paper, 4.5 x 20 in., on each of which are printed forty 5-place series of symbols, together with forty 5-place blank squares. Stop-watch.

Preliminaries.—Glue the 3 x 7 card of circles to the 7 x 14 cardboard, at the bottom. Glue the narrow strips of cardboard, two strips thick, along either edge of the under surface of the same cardboard, so as to form a 'tunnel' underneath the cardboard, through which the test-strips may slide.

Method.—Lay the prepared cardboard flat upon the table. Insert a test-strip, so that the first line of characters comes just

below the bottom edge of the cardboard, and hence just beneath the nine circles.

Cover the test-material, and do not allow *S* to examine it (save as specified below) before the test.

Give *S* the following explanation: "You will find on the table a sheet of cardboard, with nine circles at the bottom of it. In each circle you will find a number, running from 1 to 9, and a symbol, *i. e.*, a small character or drawing. Then, you will find a strip of paper with rows of the same characters, and with empty squares beside them. What you are to do is to write in these empty squares the numbers that correspond with the characters. Keep at work continuously, as fast as you can, until you have filled in all the empty squares on the paper. Of course, you will have to look back and forth from the paper to the circles to find out what number to use, unless you can, after a while, remember some of the numbers without looking at them."

With young *S*'s, this verbal explanation will be insufficient to make the task clear. It will do no harm, in such cases, to show *S*, for a brief instant, the card of circles, and a test-strip that has already been filled out. Let him see them just long enough to make the instructions clear, but not long enough to permit him to learn any of the associations.

Start the watch when *S* starts the first line: keep the watch in view, but out of *S*'s sight: record, without stopping the watch, the position of the second-hand when *S* completes every 5th line (indicated, for this purpose, by a heavier division-line in the test-strip.)

As fast as *S* finishes a line (or two lines), let him push the strip forward to bring a fresh line of symbols into position at the edge of the pattern (card of circles).

When the 40th line is written, conceal the circles, immediately turn over the test-strip, write on it the digits 1 to 9, and ask *S* to place above each digit the character that accompanies it. Ascertain, if possible, whether *S* relied upon visual, auditory, visual-auditory, or some other type of associative imagery.

TREATMENT OF DATA.—Check up the test-strip for errors. Compare *S*'s with respect to (1) their time for the whole test, (2) their gain in the last, as related to their speed in the first 5-line section, (3) their accuracy, and (4) their knowledge of the symbols (credit-

ing 1 for each symbol correctly reproduced, and 1 for each pair of transposed symbols). Plot graphs showing the variation in speed for the eight sections.

B. FORM FOR GROUP TESTS, OR FOR SUPPLEMENTARY INDIVIDUAL TESTS

MATERIALS.—Printed form, at the top of which are shown 9 circles, as in Form A (save that different symbols are used), and in the body of which is provided, in two columns, a series of forty 5-place numbers and forty 5-place blank squares in which the appropriate symbols are to be placed. Stop-watch.

METHOD.—For individual tests, give instructions similar to those for Form A, with such modifications as the altered arrangement of the material necessitates. Make clear, especially, that the second column is to be filled out the moment that the first is completed.

For group tests, supplement the instructions by an adequate blackboard explanation, preferably with an illustration so devised as not to give information concerning the symbols to be used. Have the papers distributed, face down, to be turned over only at the command to start. Work by the time-limit method, allowing 4 min. for the test. Instruct *S*'s to place an oblique mark at the point reached when the command "mark" is heard. Give this signal every 30 sec., so that the work is divided into 8 periods of 30 sec. each. Conclude with the symbol-test as in the individual method. Plot curves for 30 sec. intervals.

VARIATIONS OF METHOD.—Cut off the top of the paper: glue the pattern of circles on a sheet of cardboard, as in Form A. Cut and paste the two test-columns to form a single long column, as in Form A. This will permit check-tests, comparable with the standard method, save that here symbols, there digits are written.

Repeat either Form A or Form B after an interval of several hours, days, or weeks, to compare the permanence, in different *S*'s, of the associative connections established in a single trial.

Repeat Form A until the associations are firmly established, and the digits can be written rapidly without seeing the pattern. Ascertain whether the use of Form B will then develop interference of associations.

Results.—(1) *Individual differences* are easily demonstrated. Thus, in a 4-min. test of 40 college students, in the same class (Table 61), the fastest *S* wrote 53 per cent more symbols than the slowest *S*, and in individual tests of 18 students and college instructors (Table 62), the fastest *S* finished in 44 per cent quicker time than the slowest *S*.

(2) *Sex differences* are less evident than in the mirror-drawing test of learning, but the evidence so far adduced is in favor of an appreciable superiority of women (Tables 61 and 62). The fastest

TABLE 61

Substitution Test. Number of Symbols Written. Form B. Group Method (Whipple)

THIRTY SEC. PERIOD	1ST	2D	3D	4TH	5TH	6TH	7TH	8TH	TOTAL	SYMBOL SCORE
Average, 12 men..	13.7	16.1	14.6	16.3	14.8	17.2	16.7	17.9	127.3	8
Average, 28 women	13.9	15.4	16.0	17.9	16.0	17.0	16.8	19.0	132.0	8.2
Fastest individual	10.0	21.0	22.0	18.0	23.0	20.0	25.0	26.0	165.0	9
Slowest individual	11.0	13.0	8.0	10.0	12.0	10.0	11.0	13.0	95.0	3

TABLE 62

Substitution Test. Speed in Seconds. Form B. Individual Method (Whipple)

SECTION OF 5 LINES	1ST	2D	3D	4TH	5TH	6TH	7TH	8TH	TOTAL
Average, 13 men.......	54.0	46.0	45.8	44.8	46.1	44.4	47.7	44.3	353.1
Average, 5 women.....	45.8	41.2	40.6	38.6	43.4	37.6	36.6	35.0	318.8
Total, 18 cases..........	51.8	44.7	44.3	43.1	45.4	42.5	44.6	41.7	358.1
Fastest individual.....	42.0	35.0	33.0	30.0	36.0	29.0	31.0	34.0	270.0
Slowest individual.....	63.0	58.0	59.0	61.0	62.0	53.0	60.0	65.0	481.0

S in each group tested was a woman, the slowest *S* in each group was a man.

(3) The effect of *age* upon the outcome of the substitution test may be roughly estimated by comparing the averages for adults, in Tables 61 and 62, with the averages for grammar-school boys, in Table 63.

(4) A tendency toward a *correlation with intelligence* is indicated by the comparison of 6 dull, and 5 bright grammar-school boys

(Table 63). All the bright boys, with a single exception, are younger than the dull boys, yet they show unquestioned superiority.

(5) *Improvement curve.* (*a*) Comparison of the 8 sections into which the test has been divided shows that, in general, the formation of associative connections brings about a progressive increase in speed of work, amounting, roughly, to from 10 to 20 per cent, when the first is compared with the last section.

TABLE 63

Substitution Test. Bright and Dull Boys. Form B. Individual Method (Whipple)

	BOY	AGE	SCHOOL GRADE	TIME IN SEC.	SYMBOL SCORE
Dull Group	H.	16:9	7, II	700	6
	K.	13:1	5, II	742	8
	N.	14:9	6, I	422	9
	M.	12:8	6, I	975	1
	B.	12:6	7, II	707	4
	S.	15:2	6, I	660	9
Average....		14:2		701	6.17
Bright Group	Br.	11:11	8, I	677	9
	Hu.	12:8	6, II	597	5
	Id.	10:9	6, II	566	9
	Tr.	10:4	6, II	648	4
	Fe.	10:8	6, II	591	9
Average ...		11:1		615.8	7.2

(*b*) This increase of speed is, however, not uniform. In particular, a tendency may be discovered, *e.g.*, in Tables 61 and 62, toward a decrease of efficiency at about the middle of the work. Thus, in the individual tests, both men and women, taken collectively, exhibit a reduction of speed in the 5th section, while in the group tests, both men and women write fewer symbols in the 5th, than in the 4th 30-sec. period.

In the individual tests, the 4th section comes at the bottom of the first column, the 5th at the top of the second column. The brief delay occasioned by the necessary readjustment (of paper, pencil, attention, etc.) may explain a part, but only a part of the reduction in time.

A plausible explanation is that reported by one *S*, who noted that, in Section 4, being so far from the circles, she relied upon her memory, wherein Section 5, the very proximity of the circles tempted her to glance at them to make sure of her work, and thus to work more slowly.[1]

To test this hypothesis, trials were made with 10 college students, using material of Form B, but rearranged (as suggested above) to resemble Form A (the test-blank in one long column sliding beneath the cardboard). The average scores, in sec., per 5-line section, were 55.7, 48.6, 44.0, 40.9, *43.0*, 40.3, 41.5, and 40.8, for the sections 1 to 8, respectively. (Total time, 354.8 sec.; symbol score, 8.2). There is, then, still a loss of more than 2 sec. at Section 5.

It would appear, therefore, either that the test-material of Section 5 happens to be more difficult than that of Sections 4 and 6, or that, as a final possibility, the slower rate in Section 5 is merely an expression of a mental condition—fatigue, weariness, loss of initial enthusiasm. That this explanation may be entertained is shown in Table 64, where it will be seen that, although more *S*'s lose speed in the 5th, than in any other section, there are, nevertheless, numerous instances of loss of speed in other portions of the work, especially in Section 7. The *S*'s of Table 64 are the 10 just mentioned, and the 18 of Table 62.

TABLE 64

Substitution Test. Distribution of Gains and Losses in Speed (Whipple)

SECTIONS	1–2	2–3	3–4	4–5	5–6	6–7	7–8
Number gaining speed.......	24	17	16	9	19	12	16
Number losing speed.........	4	7	8	16	8	15	8
Number maintaining speed...	0	4	4	3	1	1	4

(6) *S*'s who make the fastest records commonly employ the scheme of holding in mind the entire 5-place number (in Form B), and writing down the symbols while keeping the eyes directed upon the circles. The material in Form A lends itself less easily to this scheme.

[1]Form A has been devised especially to avoid the variation in distance of test-blanks from the pattern at different periods of the work.

TEST 38

Memory for serial impressions: 'Rote' memory.[1] The essential idea of the several forms of memory test treated under this title is to present a series of discrete impressions, *e.g.*, letters, digits, words, which is, if possible, to be reproduced in correct order and exactly as presented. These tests are to be contrasted with the so-called tests of 'logical' memory, in which the material presented is a logically connected whole, and in which the requirement is to reproduce the substance, or the meaning, of what has been presented. In either test, the reproduction may be immediate or delayed, and the mode of presentation and method of measuring efficiency may be varied in many ways.

Memory for a series of discrete impressions has been used to study individual differences, as conditioned by sex, age, mental ability, to detect fatigue, to investigate the nature of practise, the possibility of training retention and recall, the most economical methods of learning, etc.

To understand the results and conclusions of the small army of investigators of memory, it is convenient to classify the methods and the materials that have been most commonly used.[2]

CLASSIFICATION OF METHODS FOR MEMORY TESTS

(1) The *method of complete memorization*, or method of complete mastery (*Erlernungsmethode*), developed in the classic work of Ebbinghaus (*Ueber das Gedächtnis*) in 1885, and refined by Müller and Schumann, demands that *S* repeat the series of impressions again and again until he can reproduce it without error, without hesitation, and with certainty of correctness. Efficiency is measured by the number of presentations required for this complete learning.

[1] The author desires to acknowledge the assistance of Dr. L. R. Geissler, of Cornell University, in the collation and sifting of the literature bearing upon this test.

[2] For more extended discussion of the historical development of the several experimental methods, together with accounts of the results that have been attained, the reader should consult Bentley (1), Binet (6), Burnham (11), Ebbinghaus (15b), Gamble (19), Henri (21), Offner (37), and Pohlmann (38). The last-named gives a particularly valuable summary of the methods.

In practise, this method is frequently supplemented by testing the number of presentations of the same series that is needed to relearn it at any assigned time after the first learning (*Ersparnisverfahren* or *Ersparnismethode*), in which case the saving in number of repetitions in the relearning, as compared with the learning, measures the amount of retention, or the degree to which the first impression has persisted.

(2) The *memory span method* (*Methode der Gedächtnis-Spanne*), first devised by Jacobs (22), elaborated by Ebert and Meumann, and extensively used in England and America, consists in the determination of the maximal length of a series of impressions that can be reproduced with a given degree of accuracy (usually complete accuracy) after a given number of repetitions (usually, though not necessarily, one repetition). Ordinarily, *E* begins with a series that is easily within *S*'s limit, and increases the length of the series, keeping other factors constant, until errors appear.

(3) The *method of retained members* (*Methode der behaltenen Glieder*), first so designated by Ebbinghaus, but more carefully studied by Pohlmann, consists in the determination of the degree of mastery (proportion of elements correctly reproduced) of a series of a given length, after a given number of repetitions. The method is somewhat like the span method, but the length of the series is so chosen that *S* cannot attain complete mastery. In practise, many span tests actually become tests of degree of mastery.

(4) The *method of right associates* (*Treffermethode*), proposed by Jost (24), and developed by Müller and Pilzecker (33), consists in presenting a series of impressions (typically, nonsense syllables in trochaic rhythm), and of subsequently testing *S*'s ability to name the member that follows any given member. Usually the accented member is given, and *S* tries to designate the 'right associate' for it. (When his time of response is measured, the method is known, in full, as the *Treffer- und Zeitmethode*). Its special value is to afford opportunity for analyzing the nature of the associative connections, and it has not been proposed as a test of efficiency.

(5) The *method of prompting* (*Methode der Hilfen*), somewhat similarly, tests the nature and strength of the individual associa-

tive connections in the series, and is of questionable usefulness for practical testing. As illustrated in the work of Ephrussi (17), the method consists in an attempt by *S* to reproduce the series before it has been fully learned, and in promptings by *E* at each point of hesitation or error. Efficiency is inversely related to the number of promptings required.

(6) The *method of interference of associations* is exemplified in Bergström's study of card-sorting (2): here 80 cards are sorted by *E* into 10 piles, and subsequently, at a given interval, into another 10 piles differently arranged; the second sorting is slower because of the persistence of associative connections developed in the first trial. Analogous tests can be fashioned with other forms of material, as has been suggested in the Substitution Test.

(7) The *method of reconstruction*, used by Münsterberg and Bigham with colors, and by Miss Gamble with odors (19), consists in presenting a series of stimuli in a definite order, and then, after a predetermined interval, in presenting the same stimuli in chance order. *S* attempts to rearrange them in the original order.

(8) The *method of recognition* consists in the presentation of a limited number of impressions, which are subsequently presented again, in conjunction with other stimuli, to see how many of the first series *S* can recognize in the second series. Examples will be found in the works of Smith and of Henri.

(9) The *method of identical series*, as employed by Reuther (39), is a modification of the method of recognition, in which the original series is always actually presented intact, though, of course, this fact is concealed from *S*.

(10) The *method of continuous lists* (*Methode des fortlaufenden Niederschreibens oder Aufzählens*), employed by Kræpelin (27), is identical with the procedure described in Test 33, though sometimes *S* is required to write words that belong to specified categories.

(11) The *method of chance verbal reactions* (*Methode der zufälligen Wortreactionen*), well illustrated by the investigations of Aschaffenburg and G. E. Müller, is the stock association experiment, with emphasis upon the qualitative as well as the quantitative study of the associative sequences.

(12) The *method of description or report* (*Aussage*) is a form of memory investigation with peculiar problems of its own, as has

been shown in Test 32. In it, the terms in which the reproduction takes place are not restricted to a direct equivalence with the material presented, but are merely indicative or descriptive of this material.

CLASSIFICATION OF MATERIAL FOR MEMORY TESTS

The material used in tests of serial memory may be classed according to the sense-department to which it is presented (visual, auditory, visual-auditory, etc.), and according to its nature or form. Again, visual material of different forms may be presented either simultaneously or successively.

(1) *Actual objects* were used by Netschajeff (35), Lobsien (30) and Kirkpatrick (26). Thus, Lobsien showed 9 objects at the rate of 1 per sec., *e.g.*, newspaper, key, handkerchief, glass, slate, box, book, glove, chalk. Netschajeff used 12, Kirkpatrick 10 objects.

(2) *Pictures of objects*, 10 in number, were used by Miss Calkins (13).

(3) *Sentences* were used by Ritter (40), Miss Sharp (42), and Binet. In general they prove unsatisfactory: it is difficult to secure series that are equivalent in length, construction, difficulty, and interest. A graded test series will, however, be found in connection with the Binet-Simon tests of Chapter XIII.

(4) *Words* may be used in the most varied kinds of series. Thus, series of Latin-German, or English-German, or other pairs of nouns, have been used to produce a 'vocabulary' form of test, as by Wessely (49). A distinction may be made between 'related' or 'associable' terms and 'unrelated' or 'dissociable' terms (Norsworthy, 36: Bergström, 3). For example, *paper, writing, compose*, etc., *vs. horse, bricks, soldier, acorns*, etc. Meumann (31) and Burt (12) have compared the span (3 to 8-term series) for concrete nouns, *e.g., stove, ink, lamp, street*, etc., with the span for abstract nouns, *e.g., humanity, arrangement, organism, influence*, etc. Netschajeff and Lobsien tested the relative reproducibility of words (12 and 9-term series) that connoted visual, auditory, tactual, and emotional ideas, respectively. (Examples: *lightning, dial, sunbeam: thunder, crash, whistle: cold, soft, smooth: hope, doubt, regret.*) Kirkpatrick and Calkins also used 10-term series of words that related to objects, as did Pohlmann. Hawkins (20) compared simultaneous and successive exposure of 15 nouns. Binet, Ritter (40) and Sharp also employed lists of words of varied length and complexity.

(5) *Nonsense syllables* were tried but discarded by Jacobs. They formed, however, the stock material in Ebbinghaus' pioneer work, and were subsequently made more serviceable by the precise rules that Müller and Schu-

mann (34) formulated for their construction. Bergström, Burt, Smith, Müller and Pilzecker (33), Pohlmann (38), van Biervliet (48), and others have found them of value: indeed, Pohlmann contends that, on account of their equivalence one to another and their relative freedom from varying associations in different *S*'s, nonsense syllables form the best and most reliable material for memory tests. Series specially adapted for English readers will be found in Test 25.

(6) *Letters* (usually consonants only, to avoid the formation of syllables or words) have been used by Jacobs, Binet (5), Cohn (14), Pohlmann, Sharp (42), Smith (44, 46), and Winch (50). An idea of the great variety of procedure that may be developed with a single form of material may be gained by noting that Binet used 15 consonants exposed visually and simultaneously, for 20 sec.; Cohn exposed 12 consonants arranged in the form of a square for 25 sec.; Pohlmann read 10 consonants to his *S*'s 3 times over; Sharp exposed 12 letters successively with the Jastrow drop-apparatus, at the rate of 1 per sec., and repeated until the series was learned; Smith exposed 12 consonants simultaneously for 10 sec., and read other series of 4, 5, 6, 7, and 9-term consonants; Winch repeated 12 consonants auditorily in 25 sec., and also exposed 12 simultaneously for 25 sec.

(7) *Two-place numbers,* administered orally, were used by Schuyten (8 numbers repeated by *S*'s in concert), Lobsien (9 numbers), Pohlmann (10 numbers, given three times), and Netschajeff (12 numbers).

(8) *Digits,*[1] *i.e.,* one-place numbers, have been employed by Jacobs, Johnson (23), Bolton (9), Binet, Ebbinghaus (15a), Hawkins (20), Ritter, Sharp, Smedley (43), Krueger and Spearman (28), Wissler (52), and many others, in the most varied manner (4 to 10-place series, given auditorily, visually—either simultaneously or successively—or in combined appeal to vision and audition, to vision, audition, and 'hand' memory, or to vision, audition, and 'articulatory' memory).

(9) *Geometrical drawings* have been used by Münsterberg and Bigham, and by Bernstein and Bogdanoff (4), the latter selecting forms that would be unfamiliar to their *S*'s.

(10) Lines of varied lengths have been employed by Toulouse and by Binet (6).

(11) *Miscellaneous visual characters,* symbols, combinations of dots, lines, etc., formed a portion of the material in the investigation of Ebert and Meumann.

---

[1] Reuther (39) has formulated rules for the construction of test-series of digits, analogous to the rules of Müller and Schumann for test-series of nonsense syllables. The following are the most important of Reuther's principles: (1) Do not repeat a digit in the same series (impossible to avoid, of course, in 10-place series). (2) Do not begin a series with the number 1. (3) Avoid the use of zero. (4) Do not place any two digits in their natural relations with one another. (5) Do not use sequences that suggest historical dates. (6) Do not use in immediate succession two series that have the same digit in the same place at any point in the series.

(12) *Sounds*, such as those produced by tearing paper, whistling, stamping, ringing a bell, etc., were arranged in 9-element series by Lobsien, and in 12-element series by Netschajeff.

Aside from these wide differences in general method and in form of material, attention should be called to differences in rate or tempo at which the series is first presented, to differences in the number of times the series is presented, and to differences in the time-interval elapsing between presentation and reproduction.

As a rule, the rate of presentation has been not slower than 1 impression in 2 sec., and not faster than 2 impressions in 1 sec. A rate of 1 impression in 0.75 sec. has been found well adapted for adults.

The typical span test is one in which the series is presented but once: from the point of view of functional testing, therefore, the repetition of the stimulus series may be regarded as a variant method, not to be introduced save for the special purpose of studying its effect.

Similarly, the greater portion of the tests here reported have been made with no interval between presentation and reproduction. It is to be noted, however, that Smedley, in his tests of Chicago school children, separated presentation and reproduction by an interval of 5 sec. Kirkpatrick, and Calkins in her repetition of his tests, secured a reproduction both immediately after, and 3 days after the presentation, in order to contrast 'immediate' with 'delayed' memory or recall. Somewhat similarly, Binet, and Sharp in her repetition of his tests, secured a reproduction of each of seven 7-place word-lists directly after its presentation, and a 'recapitulation,' in so far as it was possible, of the 49 terms at the close of the whole test, *i.e.*, about 3 min. after the first presentation. Binet contrasts, in this way, immediate memory with what he terms 'memory of conservation.'

Since, as the results that follow show, even minor variations in the conduct of a memory test affect its outcome, it follows that the results of different investigators may not be expected to exhibit complete accordance with respect to the relative influence of sex, age, mental ability, etc.

Three chief forms of test have been selected and are recommended as standard for this field of investigation: variant methods are suggested in each case; but, by reference to the classification of methods and materials just given, *E* can devise further modifications to suit special requirements. These three forms are (1) tests with digits, resembling in scope Smedley's Chicago tests, but with several differences in procedure, (2) tests with letters, after Cohn's method, and (3) tests with lists of words, after the methods of Meumann and of Burt.

A. IMMEDIATE MEMORY TESTS WITH DIGITS

Materials.—Printed test-cards, 42 in number, arranged in three sets of 14 cards each, for presentation by 3 different methods. Each set contains 2 cards each of 4, 5, 6, 7, 8, 9, and 10 digits. Metronome (Fig. 14). [For serial visual exposure, in addition, Jastrow's memory apparatus (Fig. 54). Cardboard. Willson's gummed figures, black, Size 5. For letter tests, full sets of gummed letters, Sizes 5 and 10.]

Preliminaries.—On the back of each card, write the digits that are printed on its face: this enables *E*, when the test demands it, to pronounce the test numbers while displaying the card to *S*. The purely auditory and the auditory-visual-hand-motor series are not included in the printed cards, but should be prepared by *E*, preferably, for convenience, on a single piece of cardboard, the size of the printed cards. For the auditory series, use the following numbers, in the order given: 6135, 2947, 36814, 57296, 241637, 935816, 8537142, 9412837, 47293815, 71836245, 924738615, 475296318, 8697132504, 2146073859. For the visual-auditory-hand-motor series, use these numbers, reversed, *e.g.*, 5316, etc.

Method.—If only a single test can be made, employ the visual-auditory-articulatory form of presentation, since this is most likely to produce uniform conditions of ideational imagery for all *S*'s. But if the tests can be taken in full, follow the order of presentation outlined herewith.[1] In any event, preface each form of presentation with a special, short sample-series, without demanding reproduction, in order that *S* may be perfectly clear as to the nature of the test. Within each form of test, also, preface each presentation with a statement of the number of members in the coming series, *e.g.*: "This will be a series of 5 digits." The metronome should be set at 60, *i.e.*, one stroke per sec., for all tests.[2]

[1] It goes without explanation that the longer series may be omitted with very young, the shorter with mature *S*'s. Use, for the shortest series, one that is easily within the span of the poorest *S* to be tested, for the longest series, one that is too difficult for the best *S* to reproduce without error.

[2] It may be well at this place to point out the differences between this procedure and that followed by Smedley at Chicago. Smedley used no series longer than 8. He gave no warning of the length of the coming series. He set the metronome at 90. He did not present the several series in regular order, but irregularly, though beginning with an easy series. He inserted an interval of 5 sec. between presentation and reproduction. He distributed

(1) *Auditory presentation.* Explain the test by a simple illustrative series. Require *S* to close his lips firmly, and to press his tongue against the roof of his mouth—this to reduce the tendency to articulation, and in group tests (all of the memory tests lend themselves well to group presentation) to avert communication between *S*'s. Start the metronome.[1] Pronounce the digits, one at a time, with the utmost care to ensure even tempo, clear articulation, and entire absence of rhythm.[2] Directly at the conclusion of the series, let *S* repeat as much as possible of it. Although, under some circumstances (with very young or backward *S*'s), an oral reproduction may be imperative, a written reproduction should be considered standard, both because the proper placing of the digits furnishes *E* with data for scoring *S*'s performance (and the placing must indicate possible omissions), and because experiment shows that, at least for maturer *S*'s, written reproduction is preferred, and is more successful than oral reproduction. *S*'s recall should, therefore, be entered upon a prepared blank, with the caution to indicate every omission by a dash or a blank space.[3]

(2) *Visual presentation.* Use Cards V-4a, V-4b, etc. to V-10b. Follow the directions for auditory presentation, but in place of

his tests, seven in all, at hourly intervals. Finally, he gives no clear statement of his method of computing results, save that the "percentage correctly recorded constituted the grade."

[1] If he finds it necessary, *E* may substitute a silent metronome, made by swinging a small weight on a string, but the fact that the regular metronome is somewhat noisy should not be taken as evidence that it disturbs *S*: on the contrary, a noise of moderate intensity is not infrequently found to be a stimulus to better attention. Moreover, the ticking metronome is much more serviceable when *S* is asked to pronounce the digits in conjunction with *E*.

[2] The difficulty of speaking without accent, or without grouping the digits, has led Binet to reject oral, in favor of visual presentation. Even if *E* pronounces without accent or rhythm, there is no guarantee that *S* may not mentally cast the digits into a strongly accented and grouped series, and in fact, mature *S*'s, working with the longer series, are almost certain to catch this 'trick' in time. Ritter (40) advocates that *E* should give a decided objective rhythm to every series on just this account: this factor will then form a constant, rather than a variable 'error.' One difficulty with this plan lies in the fact that, in using series of varying lengths, it is impossible to use any constant metrical phrasing.

[3] For group work, the class should be provided with blank forms, so numbered and arranged that no misunderstandings may occur on the part of *S* in entering the data, or on the part of *E* in interpreting it. Allow ample time for writing: Netschajeff, Pohlmann and Schuyten all found 2 min. desirable in classroom tests. In group tests, care must be taken to prevent audible repetition of the digits during the reproduction.

pronunciation, exhibit the entire card for a length of time identical with that for auditory presentation, *i. e.*, allowing 1 sec. per digit.[1] Note to what extent *S* articulates the digits: even with lips and tongue placed as directed, they will often be seen to move, and contractions of throat muscles may also indicate partial articulation.

(3) *Auditory-visual presentation.* *E* presents the cards, as in the purely visual procedure, but also pronounces the digits, as in the auditory procedure, by reading them from the back of the card. *S* sees and hears the digits. Cards AV-4a to AV-10b are used.

(4) *Auditory-visual-articulatory presentation.* *E* presents the cards as in (2). *E* and *S* pronounce them in concert, in time with the metronome. *S* sees, hears, and pronounces the digits. Cards AVA-4a to AVA-10b are used.

(5) *Auditory-visual-hand-motor presentation.* *E* pronounces the digits as in (1): *S* writes them, as fast as pronounced by *E*, upon scrap paper: when the series is finished, *S* at once discards the scrap paper, and reproduces the series. *S* hears, sees, and writes the digits. Use the same numbers as in (1), but reverse the order of the digits. In this test, it will ordinarily be necessary to devote one or two preliminary trials to fore-exercise.

Variations of Method.—(1) Substitute successive for simultaneous visual presentation in Forms 2, 3, and 4. For this purpose, *E* must prepare cards for insertion in the Jastrow memory apparatus,[2] so that the numbers used in Forms 2, 3, and 4 (above) may now be exposed in vertical columns. In order to secure sufficiently

[1] The metronome should be used in this part of the test to keep the conditions comparable with the other presentations: it probably, also, tends to induce *S*'s to apprehend the digits successively, in the same tempo as that used in auditory presentation. For successive presentation, see under *Variations of Method.*

[2] Jastrow's instrument is adequate if *E* is careful to make the exposures regularly, in time with the metronome; it is especially useful for group tests. If *E* desires a more accurate exposure apparatus, for individual tests, he may employ the Ranschburg memory-apparatus (now improved by Wirth), Bergström's rather elaborate exposure apparatus, or G. E. Müller's modification of the kymograph for 'step-fashion' exposure, as described, in improved form, by McDougall (32). Burt, however, contends that the distraction produced in immature and inexperienced *S*'s by the sight of unfamiliar apparatus more than counterbalances the advantage of greater precision, mechanical regulation of rate and duration of exposure, etc.: he used, for successive exposure, a slotted piece of cardboard, which was shoved along the column of impressions by *E* (apparently at no uniform or constant rate, but as fast as proved convenient to *S*).

long series, the exposure-lever of the instrument is so inserted as to articulate with the pegs that provide a drop of 1 in. at each exposure. Black letters $\frac{7}{8}$ in. high (Willson's, Size 5) may then be used. These are visible to the normal eye at 50 ft., but E should

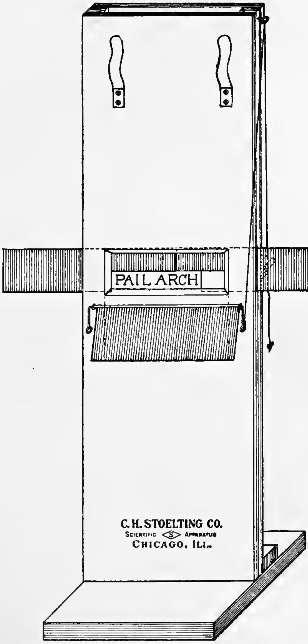


FIG. 54. JASTROW'S MEMORY APPARATUS.

take the precaution, in classroom tests, to seat myopic S's near the front of the room.

(2) Test the effect, upon a series too long for S to reproduce in one presentation, of two, three, or more presentations in immediate succession.

(3) Give repeated tests by the same method, with a series of a given length (in excess of *S*'s span), to test the effect of practise.

(4) Change the rate of exposure from one impression per sec· to one impression in 2 sec.

(5) Keeping other conditions (form of presentation, length of series, etc.) constant, compare *S*'s efficiency under normal conditions with that under different forms of distraction. Smith (44) used for this purpose three different concomitant activities: his *S*'s were required during the presentation (*a*) to tap in time with the beat of a metronome, (*b*) to repeat the syllable *la*, or (*c*) to add mentally by 2's or by 3's.

(6) Prepare cards with letters[1] in place of digits, for use by any of the procedures above described. Use only consonants. Avoid alphabetical sequences, or suggestions of words or abbreviations.

TREATMENT OF DATA.—(1) If it is desired only to determine *S*'s memory span, *sensu stricto*, this is indicated simply by the maximal number of digits that can be reproduced without error of any kind.

(2) If, as is more usual in comparative tests, it is desired to determine the degree of correctness with which series longer than the span are reproduced, the simplest plan is to assign arbitrary scores to the various forms of error. Ebbinghaus (15a), for example, scored every omission as 1 error, every displacement from the correct position in the series by 2 or 3 places as 0.5 error, and every displacement by 4 or more places as 1 error. *S*'s should then be compared with respect to their error-score in series of each length, separately.

(3) A more scientific method of determining efficiency is that of computing the degree of correlation between the order of impressions as reproduced by *S* and their order as presented. This is accomplished, following the example of Krueger and Spearman (28), by applying Spearman's 'footrule' formula for correlation (see pp. 34–35), though, in this connection, it is better to modify this formula by counting the sum of *all* the deviations between the two series, rather than the sum of all the positive, or of all of the negative deviations.

[1] Use Willson's black gummed letters, Size 5, for the Jastrow apparatus, or Size 10 to duplicate the regular printed test-cards.

For treating the data of these memory tests, therefore, the formula may be written:

$$R = 1 - \frac{\Sigma d}{(n^2 - 1)/3}.$$

The computation of Σd needs a little explanation. The following cases may be considered:[1]

(*a*) Suppose that *S* reproduces all the terms of the original series, but not in the correct order. The sum of the deviations is then easily computed. In Case A, Table 65, for instance, the sum of the deviations is 6, and, since $n = 10$, by the formula just given, $R = 0.82$.

TABLE 65

Use of the 'Footrule' Method in Scoring the Memory Test (Spearman)

ORIGINAL SERIES	CASE A		CASE B		CASE C	
	Reproduced	Deviations	Reproduced	Deviations	Reproduced	Deviations
3	3	0	3	0	3	0
7	7	0	7	0	7	0
4	2	2	–	?	2	2
2	9	1	9	1	9	1
9	4	1	4	1	4	1
1	1	0	1	0	2	?
0	0	0	–	?	0	0
8	5	1	–	?	5	1
5	8	1	8	1	8	1
6	6	0	6	0	6	0
Sum of deviations		6		15.9		9.3

(*b*) Suppose, Case B, that certain terms have been omitted. The deviations of the terms given are figured as before. There is then added the amount of deviation to be expected for the omitted terms, on the assumption that they are distributed by mere chance. The chance deviation for each term is $(n^2–1) \div 3n$. In Case B, then, there are three omitted terms, each of which deviates by chance 3.3 places. Hence, the total deviation = 6+9.9=15.9.

(*c*) Suppose that *S* reproduces certain terms more than once, *e.g.*, the digit 2 in Case C. In this case, the nearer of the two digits is considered

[1] The author is indebted for these illustrations to a personal communication from Professor Spearman.

as the correct one. The other, or duplicated, term should be regarded as an omission, and treated by the formula just given. Thus, in Case C, the total deviation = 6+3.3 = 9.3.

(*d*) Suppose that more than the correct number of terms are reproduced: here the superfluous numbers may be ignored, since, save in exceptional cases, they bring about their own penalty by disturbing the correspondence of order.

B. THE METHOD OF LETTER SQUARES

The idea of displaying simultaneously a series of consonants in a simple spatial pattern appears first to have been suggested by Binet and Henri (8): the method was extended by Cohn (14), who used it to compare the relative values, for a given *S*, of visual and of auditory-motor learning; and it has since been frequently used as a method of studying ideational type.[1] Winch used the letter square to contrast immediate with delayed reproduction, and in general to study the relation of memory to age, sex, and school standing, Smith (44) to compare the effect of various forms of distraction.

MATERIALS.—A set of 10 printed test-cards. Prepared forms upon which the reproduction is entered. Stop-watch. [The letter-square cards are printed in large type to make the test available for group procedure. The arrangement avoids the use of collocations that might serve as aids to memory. Only consonants are used. The blank forms are ruled in 12 squares.]

METHOD.—Explain to *S* the general nature of the test. Inform him of the duration of exposure, but give him no directions as to how he shall attempt to learn the arrangement of the letters. Expose the stimulus card for 25 sec. Let him fill out the blank form immediately after the exposure. Allow 20 sec. for writing. Repeat with other cards, until 4 to 10 trials have been made.

VARIATIONS OF METHOD.—(1) Defer the reproduction for 20 sec. (or 10 sec., to follow Cohn) after the exposure. Direct *S* to count aloud during this interval, from 1 to 20, 1 number per sec., in time with *E* (who may follow a silent metronome swinging once per sec.). The object is to subdue or eliminate the 'memory after-image,' and to secure true recall—in the strict sense of recall-

[1] See, for instance, the modifications introduced for this purpose, in the Cornell laboratory, as described by Titchener (47, pp. 396 ff.).

ing an experience which had not been just previously in consciousness.

(2) Direct *S* to read the letters aloud, twice over, in concert with *E*, at the rate of 1 letter per sec. Read by horizontal lines. Reproduce with, or without the 20 sec. interval.

(3) Direct *S* to repeat aloud, continuously and rapidly, during the exposure, the syllable 'Ah.' Reproduce, preferably, after the 20 sec. interval filled with the counting. This form of procedure obviously favors the visual memory. If more than one trial is made, use other syllables, such as 'La,' 'Oh,' etc., to avoid the lapse of articulation to automatism.

(4) Direct *S* to count aloud by 2's during the exposure (*e.g.*, 2, 4, 6, or 3, 5, 7, etc.) or to count backwards from 20.

(5) After exposure by any of the methods just outlined, point to one square after another, in irregular order, asking *S* to name or to write the appropriate letters as rapidly as possible. Or, without previous warning, ask *S* to fill in the blank squares in vertical rows, or in horizontal rows from right to left. In theory, visual-minded *S*'s can accomplish this without effort, whereas purely auditory-minded *S*'s must retrace their verbal associations to find the necessary letters.

Treatment of Data.—(1) Following Winch, assign 3 for each letter in its right position, 2 for each letter one remove to the right or left, or above, or below its right position, 1 for each letter two removes to the right, or left, or above, or below.

Specimen of test given				*Specimen of a marked paper*			
M	T	D	X	M(3)	T(3)	L(0)	R(1)
V	L	Y	N	L(2)	V(2)	Y(3)	N(3)
S	Z	B	R	Z(2)	B(2)	S(1)	X(1)

Score: 23 out of a possible 36.

(2) If *S* be competent to render introspective accounts of the manner in which each letter was recalled and placed, *E* may, for qualitative purposes, compute separately the score for letters recalled visually, auditorily, or in other ways.

C. MEMORY FOR CONCRETE AND FOR ABSTRACT WORDS

The essential idea of this test of memory, as devised by Meumann, and followed, with some modifications, by Burt, is to com-

pare *S*'s reproduction of a list of concrete, with his reproduction of a list of abstract terms, given under identical conditions. The comparison is based not only upon the simple quantitative efficiency in the two forms of test, but also, and more particularly, upon the qualitative analysis of the errors in the reproduced lists. Moreover, the test aims to determine not only *S*'s capacity for immediate memory, but also his degree of intelligence, or grade of mental development. The test rests in principle upon two propositions; first, that words whose meaning is understood are more easily retained and reproduced than words whose meaning is not understood; secondly, that progressive mental development implies progressive comprehension of abstract words.

Material.—For auditory presentation, use the following lists.[1] For visual-auditory presentation, use the same lists printed upon sheets of cardboard with Willson's gummed letters.

Method.—For group tests, follow Meumann's procedure. Explain the nature of the test. Provide each *S* with blanks so arranged that his reproductions may be properly recorded, the lists carefully separated, and dashes inserted for all words omitted. Before each presentation, notify the *S*'s of the number of words to be spoken. Enunciate with great care, and without grouping, at the rate of one word per sec. Instruct the *S*'s to write their

[1] The restriction of the abstract lists to words of one syllable, as was done by Burt, makes the difference in difficulty between them and the concrete lists much smaller than in the material selected by Meumann. The following abstract lists are proposed for use when it is desired to duplicate Meumann's conditions:

Four-term list	*Five-term list*	*Six-term list*	*Seven-term list*
Selection	Society	Conscience	Assumption
Analysis	Symbol	Investigation	Recognition
Explanation	Arrangement	Symptom	Origin
Character	Humanity	Formation	Influence
	Theory	Complexity	Development
		Experiment	Organism
			Value

Eight-term list
Behavior
Tendency
Interpretation
Condition
Opinion
Capacity
Profession
Connection

lists immediately after the presentation, and as rapidly as possible, without trying to 'write their very best.' Guard against interruption, intercommunication, or other possible disturbances. Give the series in order, as above, *i.e.*, 3-term concrete, 3-term abstract, 4-term concrete, etc.

Three-term lists		*Four-term lists*		*Five-term lists*	
Concrete	*Abstract*	*Concrete*	*Abstract*	*Concrete*	*Abstract*
Street	Time	Spoon	Phase	Ground	Tact
Ink	Art	Horse	Work	Pen	Scope
Lamp	Route	Chair	Truth	Clock	Proof
		Stone	Thing	Boy	Scheme
				Chalk	Form

Six-term lists		*Seven-term lists*		*Eight-term lists*	
Concrete	*Abstract*	*Concrete*	*Abstract*	*Concrete*	*Abstract*
Desk	Space	Ball	Craft	Coat	Law
Milk	Creed	Sponge	Myth	Girl	Thought
Hand	Pride	Glass	Rate	House	Plot
Card	Guile	Hat	Cause	Salt	Glee
Floor	Pledge	Fork	Style	Glove	Life
Cat	Cue	Stove	Youth	Watch	Rhythm
		Post	Mood	Box	Faith
				Mat	Mirth

VARIATIONS OF METHOD.—Follow Burt's plan of working with individuals, displaying the words successively, and directing *S* to pronounce them as displayed, in concert with *E*.[1] Lists of 4 to 8 nonsense syllables may also be included, as in his tests, for comparison with the easy and with the difficult words.

TREATMENT OF DATA.—The lists are to be examined first quantitatively, secondly qualitatively. The results for each *S* are entered upon a scheme like the following, which is transcribed, with some modifications, from Meumann.

ILLUSTRATION OF THE TREATMENT OF DATA IN MEMORY FOR WORDS

Subject: Adolph L. Age, 8 years.

Types of Error	*Number*
1. Memory errors (omissions and displacements), concrete lists.	$5\frac{2}{3}$
2. Memory errors (omissions and displacements), abstract lists	$7\frac{7}{3}$
3. Insertions	4

[1] It is to be expected that Burt's method of presentation would have produced, as it did, less characteristic differences between the two kinds of words than were found by Meumann. An unfamiliar, or but slightly familiar word has an obviously better chance to be recalled if it can be seen, heard, and pronounced, instead of being merely heard.

4.	Insertions of nonsense words	1
5.	Fusions	0
6.	Perseverations	3
7.	Regressive inhibitions	1
8.	Complete reversals	1
9.	Substitution of synonyms	0
10.	Substitution of concrete for abstract	1
11.	Wrong formations	4
12.	Misunderstood abstract terms	5
13.	Spelling	Very bad
14.	Handwriting	Undeveloped and ugly

(1) and (2) *Omissions* are represented by the integers, *i.e.*, Adolph L. omitted five words from the concrete lists, 7 from the abstract (the test was carried to the 7-term list only). *Displacements* from the correct order count $\frac{1}{3}$ error when the displacement is by one remove only, $\frac{2}{3}$ error, when more than one remove (save that with younger children, as in the case above, all displacements count $\frac{1}{3}$.) Hence Adolph L. made 2 displacements in the concrete, 7 in the abstract series.

(3) *Insertions* are the total number of words added. These are counted as 1 error each, unless the added word has some similarity of sound to a word actually presented, in which case it counts $\frac{2}{3}$ error.

(4) This rubric embraces the relatively infrequent addition of a *meaningless word* that has no similarity in sound or spelling to any of those presented.

(5) *Fusions* of two or more totally independent, successive terms into a single meaningless term are a very significant form of error, which appears in abstract lists written by *S*'s of poor intelligence, *e.g.*, *Organ* and *Gattung* are reproduced as *Orgattung*. Mostly found in children 8 and 9 years old.

(6) *Perseverations* are indicated by the recording by *S* in a given series of a word that had already been reproduced in an earlier series. If frequent, this is a sign of a low intelligence, lack of self-control and of critical judgment.

(7) *Regressive inhibitions.* Failure to reproduce at least one-half of the terms given is, as a rule, to be interpreted as regressive inhibition. This condition is commonly attributable to a state of confusion into which a child is thrown, when he is suddenly 'overwhelmed' by the task, when everything 'flies out of his mind,' he 'loses his wits,' and is unable to accomplish even a fraction of his normal performance. The same thing is seen in adults under conditions which are difficult for them. Since, Meumann argues, this is essentially due to inability to force attention, lack of this ability is a token of poor general ability, and hence of low intelligence. Failure due to absolute lack of intent to succeed must, of course, be distinguished from the lack of ability to succeed.

(8) *Complete reversal* of word order, either in a large portion, or in the

whole of a list is "a peculiarly puzzling phenomenon." There are occasionally met, for instance, cases in which a series of 8 words are all written in the reverse of the order presented.[1]

(9) The *substitution of synonyms* refers to the easily intelligible cases in which a word of like meaning, but different sound, replaces the word given, *e.g.*, *road* for *street*.

(10) The *substitution of concrete for abstract words* refers to the use of concrete terms of similar sound, whether of similar meaning or not, *e.g.*, *cows* for *cause*, *simple* for *symbol*. *E* must use his judgment here in making allowances for faulty spelling.

(11) Wrong formations, in especial the use of wrong endings, constituted a prolific source of error in the German tests, particularly with abstract words, *e.g.*, *Glaubheit* for *Glaube*. Errors of this type may be expected to be less frequent in the less highly inflected and compounded English language, but occasional instances will be found, *e.g.*, *selectness* for *selection*.

(12) *Misunderstood abstract terms* is to be regarded (as the author understands it) as expressing the sum total of misapprehended abstract terms, whether the misunderstanding is indicated by substitutions, faulty endings, fusions, very faulty misspellings, or in other ways.

(13) *Orthography* constitutes a secondary symptom of intelligence. In order to estimate spelling fairly, papers are ranked as 'poor' in spelling only when the sum of misspelled words is 50% or more greater than the average number of misspellings for *S*'s class.

(14) *Handwriting* constitutes another secondary symptom of intelligence, and is merely rated, as fairly as possible by comparison of numerous papers, as good, average, or poor.

These 14 rubrics are filled out for each *S*. For the estimation of memory capacity, pure and simple, Meumann takes Nos. 1 and 2; for the estimation of intellectual ability, he divides the rubrics into three groups, (1) those that serve as indirect indexes of intelligence (Nos. 1, 2, and 3), (2) those that serve as direct evidence of intelligence (Nos. 4 to 12, including a statement of the relation of Nos. 1 and 2), and (3) those that serve as secondary symptoms of degree of mental development (Nos. 13 and 14). Now, for each of these condensed indexes, the grade of each *S* is indicated as (1) above average, (2) average, or (3) below average, and final comparisons and correlations are based upon these grades.

[1] The author is inclined to regard this phenomenon as a simple case of attempt on the part of a few *S*'s to get the series right by beginning with the last word heard and working back to the first section. *S* may have intentionally disregarded instructions to reproduce in the order given, or may have interpreted these instructions to include the reverse order as acceptable. In other words, it scarcely seems probable that the child does not know that he has reversed the order of presentation.

TABLE 66

Norms of Memory Span for Digits, as Conditioned by Age (Smedley)

AGE	AUDITORY SPAN	VISUAL SPAN	AGE	AUDITORY SPAN	VISUAL SPAN
7	5	5	14	6	7
8	5	5	15	6	7
9	5	6	16	6	7
10	6	6	17	7	8
11	6	6	18	6	7
12	6	7	19	7	8
13	6	7			

GENERAL RESULTS AND CONCLUSIONS.–(1) *Norms of performance* for the digit test (Smedley's method) are reproduced in Tables 66 and 67, for the letter-square method (from Winch) in Table 68.

Neither Meumann nor Burt has published data for the word test in the form of norms. Typical performances for other varieties of memory test will be found in the tables that follow.

TABLE 67

Development of Memory for Digits (Smedley)

AVERAGE AGE		NUMBER TESTED	PER CENT REPRODUCED	
Years	Months		Auditory	Visual
7	8	19	36.4	35.2
8	8	58	44.6	42.8
9	6	100	45.0	47.4
10	5	89	49.4	54.6
11	6	91	55.4	64.7
12	6	93	55.7	72.3
13	7	109	57.9	76.8
14	6	114	66.2	80 5
15	6	94	65.6	78.2
16	6	77	66.9	81.3
17	6	56	65.5	84.1
18	5	25	67.2	77.5
19	5	12	70.0	85.3

(2) *Dependence on sex.* The investigations of Bolton, Kirkpatrick, Calkins, Schuyten (Table 69), and Pohlmann all agree in showing a superiority of girls over boys and of women over men in tests of immediate or of 'delayed' (3 days) memory. Lobsien's tests with varied materials (Table 74) likewise showed that girls reproduced more, but that boys were more apt to get the order right.[1] Netschajeff also concluded that girls made more illusory errors (especially at ages 9 to 11). He also found that boys had the better memory for real objects, girls for numbers and words, in which they surpassed boys, particularly during the years 11 to 14.

TABLE 68

Memory for Letter Squares, in Relation to Age and Practise (Winch)

SCHOOL GRADE	NUMBER TESTED	AVERAGE AGE	AVERAGE SCORE			Average for 3 Sets
			1st 10 Tests	2d 10 Tests	3d 10 Tests	
Ex-vii.	5	14 yrs. 3 mos.	23.8	29.0	31.7	28.1
vii.	5	13 " 5 "	26.3	27.9	31.1	28.4
vi.	5	12 " 3 "	26.8	32.0	34.6	31.1[2]
v.	5	11 " 4 "	18.4	22.9	26.3	22.5
iv.	6	10 " 5 "	21.3	24.8	26.6	24.2
iii.	6	9 " 0 "	14.1	17.7	19.7	17.1
ii.	6	8 " 2 "	13.2	16.8	17.2	15.7

Wissler's tabulation of the freshmen tests at Columbia University and Barnard College reveals sex differences in memory span for digits that are less than the P. E. of the averages, and that favor the men for auditory, and the women for visual series (Table 70).

TABLE 69

Percentage of Accuracy in Memory for 2-place Numbers (Schuyten)

		MORNING	AFTERNOON
First test (Afternoon first)	Boys	58.1	64.0
	Girls	69.6	77.5
Second test (Morning first)	Boys	57.9	35.0
	Girls	62.6	55.1

[1] Note analogous results in the Test of Report (No. 32, p. 306).
[2] The girls of this group proved to have been of special ability.

(3) *Dependence on age.* That memory capacity increases in general from the early to the late school years is illustrated in Tables 66, 67, 68, 71, 72, 73, 75, and 78. The general evidence is fairly clear that the improvement is steady up to puberty, but that it suffers fluctuations after that period. Several investigators adduce evidence that corroborates the popular notion that there exists a special 'memory period,' or stage of maximal efficiency somewhere in the 'teens,' when memory is stronger than it is later. For example, the very careful work of Pohlmann, with varied materials

TABLE 70

Sex Differences in Memory Span for Digits in College Freshmen (Wissler)

	AUDITORY PRESENTATION			VISUAL PRESENTATION		
	Number	Average	P. E.	Number	Average	P. E.
Men	266	7.6	0.4	142	6.9	0.5
Women	42	7.3	0.5	42	7.3	0.4

and varied forms of presentation, yields the net results (method of retained members) shown in Table 71, in which maximal efficiency is indicated at 14, followed by fluctuations, without real improvement through the adolescent period. Bourdon (10) could discern progress from 8 to 13, but not from 14 to 20. Bernstein and Bogdanoff, in testing memory for geometrical figures by the

TABLE 71

Net Efficiency of Various Memories, in Relation to Age (Pohlmann)

AGE	9	10	11	12	13	14	15	16	17	18	19	20
Average Capacity.	39.4	41.4	55.7	59.1	62.1	68.9	55.3	62.9	58.6	58.0	65.4	68.3

method of recognition, found that 23 S's aged 14 and 15 averaged better than the 55 adults that they tested. Wessely (49), who tested retention during a long period (1 and 2 years), was convinced that ability to retain and reproduce poems is maximal at the years 12–14, and that vocabularies (Latin-German) are repro-

duced more accurately at the expiration of 1 to 4 weeks, when learned by 12-year, than when learned by 15-year old *S*'s. Similar assertions concerning the relative amount of retentive capacity for poems by children and by adults are made by Larguier (29, 185 ff.), while Binet believes (5, 259 ff.) that children have the better retentive capacity, and adults the better attentive capacity.

TABLE 72

Dependence of Memory Span for Auditory Digits on Age (Jacobs)

AGE	8	9	10	11	12	13	14	15	16	17	18	19
Number tested	8	13	19	36	41	42	42	72	66	50	30	14
Average Span	6.6	6.7	6.8	7.2	7.4	7.3	7.3	7.7	8.0	8.0	8.6	8.6

Over against this evidence for a decline of efficiency after 14, we have the figures of Jacobs (Table 72) and the emphatic statement of Smedley (43, p. 49), based upon his Chicago results (Table 67),

TABLE 73

Dependence of Memory for Auditory Digits on Age (Ebbinghaus)

(Average Number of Errors per Pupil in Two Series)

AVERAGE AGE	8-DIGIT SERIES	9-DIGIT SERIES	10-DIGIT SERIES	6 TO 10 DIGITS
10.7	3.1	5.1	7.4	17.8
12.2	2.9	4.7	7.9	17.5
13.2	1.5	2.6	4.2	9.1
14.4	1.6	3.0	4.9	10.5
15.5	1.0	2.1	3.7	7.6
17.1	0.8	1.4	3.9	6.5
18.0	0.9	1.4	3.4	6.1

that "there is no 'memory period,' no period in early school life when the memory is stronger than it is at any later portion of the child's life." Smedley's records do, indeed, show that "auditory memory develops rapidly up to about 14 years of age, and but slowly after this period. The visual memory seems to develop rapidly up to about 15 or 16 years of age." . . . "It will be noted

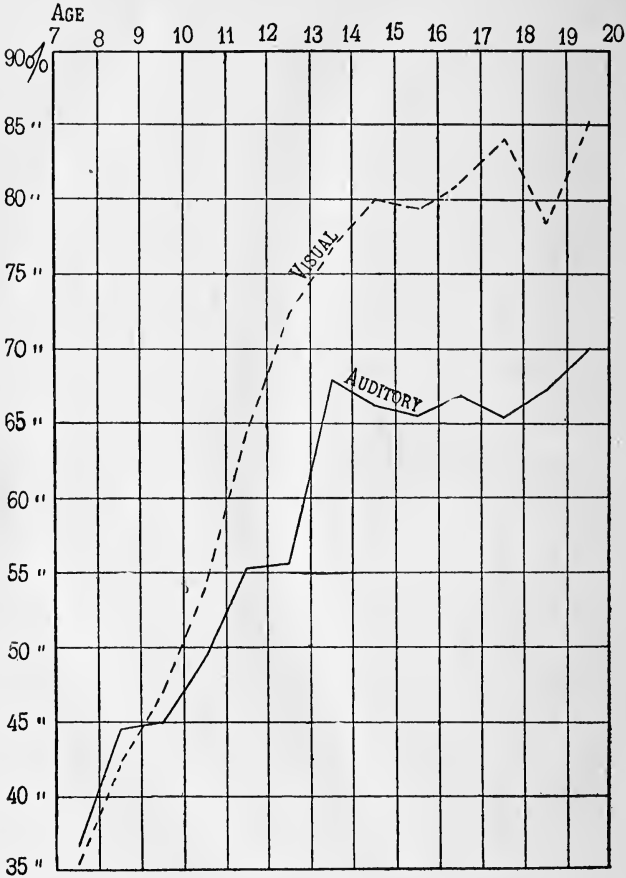


FIG. 55. DEVELOPMENT OF MEMORY FOR DIGITS.
From Smedley.

[Fig. 55] that, in the early life of the child, the auditory memory is stronger than the visual memory; after about 9 years of age, the visual memory of most of the children becomes stronger than the auditory memory, and continues to develop more rapidly than the auditory memory throughout school life. Yet, even in the high school, there still remains a small proportion of the pupils whose hearing memory is the stronger."

The dependence of different types of memory upon age has been studied especially by Netschajeff and by Lobsien. They agree substantially that, while the various forms of memory improve with age on the whole, there are periods of rapid development, followed by no improvement or even by a reduction; that while, on the whole, the greatest improvement occurs during the years 10 to 12, and development is retarded after 14, yet the different forms of memory, considered specifically, develop at different rates, and at periods that may not coincide in the two sexes. Thus, in boys, memory for objects is at first best developed, then follow, in order of chronological development, memory for visual terms, for acoustic terms, for actual sounds, for tactual terms, for numbers, for abstract terms, and finally for emotional terms. For girls, the chronological order is: visual terms, objects, sounds, numbers, abstract terms, acoustic terms, tactual terms, emotional terms. Special stress is laid upon the parallelism of development between memory for numbers and memory for abstract terms.

In Meumann's word-list tests, those types of error that indicate poor intelligence decreased with age, until, at 14 and 15, instances of misunderstood abstract terms were limited to about 10 per cent of his *S*'s, while meaningless fusions, meaningless insertions, and the substitution of concrete for abstract terms had nearly disappeared, and the memory for abstract terms had so increased as frequently to be superior to that for concrete terms. It follows that age must always be taken into account in the interpretation of this test, particularly in estimating intelligence by it.

(4) *Dependence on the nature of the material.* (*a*) When *digits and consonants* are given under the same conditions, digits are easier to reproduce (Jacobs, Sharp), especially during the years 8 to 13 (Bourdon). But, if 10-place series are presented auditorily, thrice, the order of excellence for recall is (1) consonants, (2)

names of objects, (3) 2-place numbers, (4) nonsense syllables (Pohlmann).

(*b*) Netschajeff, Lobsien, Pohlmann, and less elaborately Kirkpatrick and Calkins, have compared memory for series made up of real objects, of numbers, of sounds, and of words having characteristically visual, auditory, tactual, or emotional associative meanings. Table 74 gives illustrative results from Lobsien. Pohlmann, however, concluded that the assumption of Netschajeff and Lobsien that the presentation of visual, auditory, and other terms arouses the visual, auditory, and other imagery that their meaning implies, is erroneous, so that the results of these investigations are of little real significance.

TABLE 74.

Memory for 9-term Series of Different Kinds (Lobsien)

KIND OF SERIES	SCORE IN PER CENT CORRECT	
	BOYS	GIRLS
Real objects	82.2	91.4
Auditory numbers	64.8	71.8
Sounds	59.6	62.2
Tactual terms	64.2	71.0
Visual terms	60.6	67.2
Auditory terms	59.4	60.2
Emotional terms	31.2	59.4
Foreign terms	24.0	23.8

Kirkpatrick, and after him Miss Calkins, found, like Netschajeff and Lobsien, that memory for objects (or pictures of objects) was superior to that for words, both for immediate and for delayed reproduction; in the latter, for example, there were recalled seven times as many objects as words. The same investigators determined the order of excellence for recall of different kinds of words to be:—visual terms, auditory terms, names of objects.

(*c*) Up to the 12th year, *concrete words* are reproduced better than abstract words, but 14 and 15-year old *S*'s frequently make better records with the latter (Meumann).

(*d*) *Related terms*, *i.e.*, a series of words not in a sentence,

but readily associated with one another, are more easily recalled than unrelated words. For data, see Table 75 from Miss Norsworthy.

(*e*) Material so arranged as to aid *localization* is more easily remembered, especially by children. For example, 12 consonants in the letter-square form are easier to recall than 12 consonants in a single line; similarly, digits pronounced in rhythm are easier to recall than digits pronounced in even tempo (Müller and Schumann). Pohlmann found grouped series to be the easier in 133 of 144 trials.

TABLE 75

Norms for Memory of Related and of Unrelated Words (Norsworthy)

Age	Related Words, 288 Cases				Unrelated Words, 270 Cases			
	Boys		Girls		Boys		Girls	
	Median	P. E.	Median	P. E.	Median	P. E.	Median	P. E.
8	13.0	1.0	13.0	1.6	11.1	1.6	11.5	1.3
9	14.0	2.0	14.0	1.7	12.2	1.7	12.4	1.4
10	15.0	1.7	15.3	1.9	12.2	1.7	14.4	1.4
11	15.0	1.7	16.5	1.7	12.5	1.8	14.3	1.4
12	16.4	1.8	16.0	1.6	12.8	1.8	14.0	1.5
13	16.5	1.8	17.0	1.5	13.5	2.1	13.5	1.5
14	16.9	1.3	17.5	1.5	13.7	2.2	14.0	1.5
15	16.0	1.3	17.5	1.5	13.7	2.2	14.0	1.5
16	17.0	1.3	17.8	1.5	14.0	2.2	14.5	1.5
Adults	16.5	1.5	17.0	1.9	12.8	1.2	13.0	1.4

(5) *Dependence on sense-department directly stimulated.* It is evident that a complete isolation of the different modalities can not be accomplished by different forms of presentation: *e. g.*, auditory-minded *S*'s may actually retain and reproduce impressions presented to the eye in auditory, or mainly in auditory terms, and so on. Ten nouns heard are, by young *S*'s, recalled better than the same number of nouns successively seen, but above 15 years the reverse is true (Hawkins).[1] More extensive comparisons have been

[1] Cf. Smedley's evidence of the superiority of auditory memory in younger *S*'s (Table 67), and Pohlmann's results (Table 76).

made by Pohlmann and by Smedley. Table 76 indicates a superiority of auditory-visual presentation over either auditory or visual presentation, alone—a result in accordance with Smedley's. Pohlmann also investigated the effect of these three forms of presentation upon numerals and nonsense syllables, with the result that for 230 *Volksschule* girls, using 10-term series, given thrice, the percentage of accuracy was, for visual-auditory 53 per cent, for visual 52 per cent, and for auditory 42 per cent, which agrees in substance, so far as it goes, with Smedley's results for digits. This investigator found the order of superiority to be: (1) auditory-visual-articulatory, (2) auditory-visual, (3) auditory-visual-hand-motor, (4 and 5) visual or auditory (depending on age). Illustrative figures for S's aged 16 years are, for the five forms just mentioned, 88.4, 86.9, 82.4, (circa) 80.0, and 66 per cent, respectively. Combined appeal is, then, most powerful, but the task of writing proves somewhat distracting.

TABLE 76

Dependence of Memory upon Form of Presentation (Pohlmann)

(Percentage of Retained Members, 10-Term Series, 350 Pupils, 9–14 Years)

NATURE OF MATERIAL	FORM OF PRESENTATION	PERCENTAGE RETAINED
1. Actual objects	Shown and named by E	$72\frac{1}{8}$
2. Actual objects	Shown, only, successively	70
3. Names of objects	Seen and heard by S	$56\frac{1}{12}$
4. Names of objects	Heard, only, by S	$55\frac{1}{6}$
5. Names of objects	Seen, only, by S	$50\frac{2}{3}$
6. Names of objects	Seen, heard, and pronounced by S	$49\frac{1}{2}$

In the upper classes, 5 becomes superior to 4.

(6) *Successive vs. simultaneous presentation.* If 15 words are exposed simultaneously or successively for equivalent lengths of time, successive presentation is easier for young, but simultaneous for older children, according to Hawkins.

(7) *Dependence on number of presentations: repetition.* Pohlmann, Lipmann, Smedley, and others have found that hearing a

series thrice or twice, instead of once, improves its recall. However, Hawkins found two hearings less effective than one or three. It is certain that more is accomplished in the first hearing than in a large number of repetitions, and that the effect of repeated presentation is different in different *S*'s, so that individual differences are more marked after many hearings than after one hearing (Smith). Smedley's test of 38 10-year pupils, with auditory digits, gave, for the first hearing 47 per cent, for the second 55 per cent, and for the third 59 per cent correct reproduction. In some of Smith's tests, 12 presentations did not double the efficiency attained in one presentation.

(8) *Dependence on rate and duration of exposure.* Bergström's tests (3) indicate that nonsense syllables exposed at the rate of one in 0.77 sec., with durations of exposure of .041, .082, .164, and .318 sec. yield practically the same results, though there is a slight preference for .82 sec.

The same investigator found that, both with auditory letter and word series and with visual nonsense-syllables series, a relatively slow rate of exposure (1.5 to 2 sec. per term) yielded more accurate results than a faster rate (one term in a fraction of a second). The slower rate is especially helpful in lists of words, and for those *S*'s that try to develop associations between the terms as they are presented. Bergström summarizes by saying: "The acquisition and retention of a series of familiar associable words varies approximately as the logarithm of the interval at which the words are spoken" (3, p. 221).

(9) *Dependence on interval between presentation and reproduction.* Relatively short intervals make, apparently, but little change in reproduction. Thus, Winch could discern no clear differences in the reproduction by school children of letter squares, with or without a 25-sec. empty interval between presentation and reproduction.

Binet and Miss Sharp compared immediate memory with 'recapitulatory' memory (memory of conservation); they both noted that the word lists in immediate reproduction seemed to be held largely by sound (so that, for example, such errors as *flower* for *floor* were common), whereas lists reproduced 3 min. later appear

to be held more often by meaning, since "the errors are usually additional words suggested from analogy of sense" (*e.g.*, *dog* suggested by *cat*, *cold* by *winter*, etc.)[1]

Attention has already been called (4, *b*, above) to the demonstration by Kirkpatrick and by Calkins that the reproducibility of different forms of material is not equally affected by a 3-day interval.

(10) *Effect of distraction.* Smith's use of the method of letter squares (44), with and without the distraction of concomitant activities, shows the order of efficiency under these conditions to be, from best to worst:—(1) without distraction, (2) with tapping to the beat of a metronome, (3) with repetition of a vowel, and (4) with counting by 2's or 3's. Cohn, with the same test, found that an auditory-motor *S* was more seriously disturbed by auditory-motor distractors than a visually minded *S*, and that, when such distraction is used, visual memory steps in to aid, provided *S*'s constitutional make-up (*Anlage*) will at all permit (14, p. 182).

(11) *Dependence on practise.* (*a*) Practise produces a measurable increase in the memory span (Bolton). In the use of nonsense syllables, indeed, the practise effect can be discerned even at the expiration of 60 days of experimental work (Müller and Schumann).

Winch, from his use of the letter square, as well as of auditory letter series, not only declares that there is a "marked and almost invariable improvement," but "that 'pure memory' is markedly improvable by practise" (50, p. 134). Thus, 38 *S*'s, ages 8 to over 14, obtained, in 3 sets of 10 tests each (1 week between the 1st and 2d, and 2 weeks between the 2d and 3d), the average scores 20.6, 24.4, and 26.6 (averages of the scores of Table 68).

(*b*) The *transfer of practise* from the specially trained form of memory to other forms of memory would appear, from theoretical grounds, to be limited to those cases in which the material, content, or form of procedure of the other forms were related to the material, content, or form of procedure of the trained form. This is essentially the conclusion reached by Ebert and Meumann (16, p. 200), who say: "The objective results of our experiment show

[1] The tendency of adults is away from rote memorizing in favor of a memory of meanings. It would, then, be interesting to see whether children exhibited these same tendencies that Sharp's university students did, or exhibited them in as marked a degree.

that special memory practise is accompanied by a general improvement of memory. This concomitant improvement does not, however, extend equally to the other 'memories,' but appears to follow the law that the specific memories participate in the improvement directly in proportion as they are related in content, or in media and method of learning to the specific memory that was trained."[1]

Winch has been led, by experiments in memorization of poetry and historical prose (51), to take the more radical stand that "improvement, gained in practise in memorizing one subject of instruction, is transferred to memory work in other subjects whose nature is certainly diverse from that in which the improvement was gained, . . . at least so far as children of these ages and attainments are concerned."

On the other hand, Wessely directly controverts the conclusions of Ebert and Meumann, and says that there seems to be no formal practise effect of memory.

(12) *Dependence on physical capacity.* Both Netschajeff and Smedley find that pupils that are larger, stronger, and better developed physically have better memories than those of the contrary type. "This suggests," says Smedley, "that the immediate sense memory is dependent upon good brain formation and nutrition." (See 43, pp. 58–59, for numerical evidence.)

(13) *Dependence on fatigue.* Though fatigue may affect immediate memory and undoubtedly does so when severe, it is difficult, if not impossible, to arrange memory tests to serve as an index of the degree of fatigue. On this point, we have the unanimous verdict of Bolton, Ebbinghaus, Schuyten, and Smedley. The prime difficulty lies in the fact that either practise or ennui affects the results more than fatigue. Ritter gave up the determination of fatigue by span tests with numerals, but he did achieve results which he considers of special value by the use of 6-term series of two-syllabled nouns. With this material, he finds that errors increase with fatigue, and he goes so far as to assert that this test is the best one available for the investigation of fatigue.

[1] Consult Ebert and Meumann for further discussion of the ultimate nature of the transfer effect.

(14) *Correlation with mental ability.* Conclusions as to the relation between memory capacity and general intelligence range from a flat denial of such relationship, as by Bolton and by Ebbinghaus[1] (Table 77), to assertions like that of Jacobs that there is a "notable concomitance" between school standing and "span of prehension." A considerable number of investigators, among whom should be cited Binet, Bourdon, Burt, Pohlmann, Smedley, Winch, and Wessely, have found at least a fairly good degree of correspondence, but have expressed their conclusions with certain restrictions or explanations that merit our consideration.

TABLE 77

Relation of Memory of Auditory Digits and Intelligence (Ebbinghaus)

	AVERAGE NUMBER OF ERRORS PER PUPIL		
	Bright Group	Average Group	Dull Group
9-digit test	84	87	84
10-digit test	147	147	135
6-10 digit tests, collectively	318	319	303

Binet (5) contrasted 6 dull and 5 bright boys, and found that, on the whole, the latter surpassed the former in memory: the difference, as in his tests of other traits, was, however, more evident at the first, than at any subsequent trial.

Burt (12) estimated intelligence in four ways and measured memory for concrete words, abstract words, and nonsense syllables. The correlation between intelligence and memory proved to be .60 for Elementary School boys, and .82 for Preparatory School boys. Burt, however, did not substantiate Meumann's results (see below) as to the relative superiority in bright children of memory for abstract words.

Pohlmann, like Binet, dealt with contrasted groups. He concludes that, while in general the better pupils have better memories, there are numerous exceptions, particularly in that poor pupils may do as well as bright pupils in the memory tests.

[1] From the data of Table 77, Ebbinghaus concludes (15a, p. 430): "That sort of primary memory power that is involved in the immediate and exact reproduction of a series of relatively simple impressions is, accordingly, not more strongly, but, if anything, slightly more weakly developed in better than in poorer intelligences." Almost the only similar conclusion is that of Wissler, who found a correlation of but 0.16 between the class standing and memory capacity of 121 Columbia freshmen.

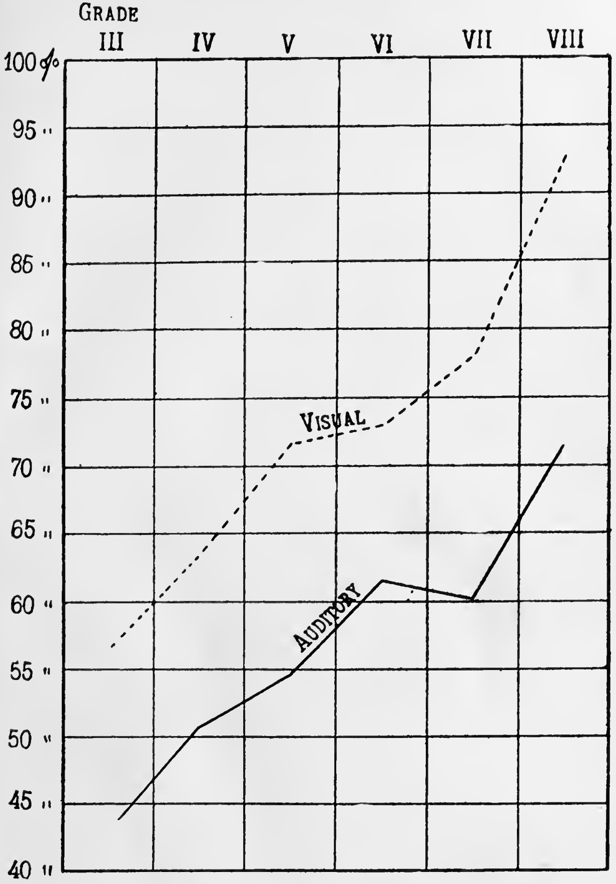


FIG. 56. MEMORY CAPACITY OF 12-YEAR PUPILS, BY GRADES.
(From Smedley.)

Smedley declares that the "parallelism between school standing and memory power holds good throughout school life" (43, p. 54), and demonstrates this by reference to mass results distributed to show the memory capacity of pupils of a given age in different grades (Fig. 56), or the capacity of pupils at and above grade as compared with the capacity of pupils below grade at different ages (Table 78).

Winch's letter-square tests convince him that "general mental ability [rank in examinations in reading, arithmetic, dictation, and English composition] is accompanied by 'good memory.'" "With two exceptions, no girl whose memory mark is relatively low has a high place in class." "'Good memory,' though usually accompanied by general efficiency, is not

TABLE 78

Relation of Memory for Digits and School Standing (Smedley)

Age	Number Tested	AUDITORY		VISUAL	
		Average Standing of Pupils At and Above Grade	Average Standing of Pupils Below Grade	Average Standing of Pupils At and Above Grade	Average Standing of Pupils Below Grade
9	99	47.8	39.7	50.3	41.9
10	88	54.4	42.7	61.6	46.2
11	91	59.0	48.6	69.4	53.3
12	92	62.6	52.2	76.7	66.0
13	110	70.4	64.3	80.7	72.3
14	116	68.9	62.6	87.6	74.9
15	94	68.9	62.4	80.9	75.0
16	75	70.1	65.8	83.3	78.8
17	56	67.5	62.7	87.8	81.2

invariably so." Again, Winch contrasted six 13-year old girls, who stood between Number 1 and Number 11 in a class of 35, with 6 girls of the same age, who stood 25th to 30th in a class of 30, and found the average score of the bright girls to be 26.9, as compared with an average score of 19 for the dull girls (50, p. 133).

Wessely believes that the correlation between memory and class standing is more evident in lower than in higher grades—a view which, if confirmed, might be explicable by the tendency to put a premium upon memorization in the lower grades.

Meumann says that the quantity of material reproduced is not in itself a reliable index of intelligence, yet the average results of mass experiments will always show that the more intelligent *S*'s have the better memory efficiency. His own experiments, he declares, were so extensive and so carefully executed as to leave no doubt at all upon this point (31, p. 78). More reliable, however, are the qualitative results attained from memory tests of the form used in his own experiments. Here, he says, virtually com-

plete coincidence is found between the several indexes of intelligence, and between them and the school marks and the estimate of mental ability by teachers. Certain characteristic indexes of poor intelligence, however, such as the fusion of abstract terms into meaningless collocations, may not be shown by all of the stupid children; if they are shown, they form a reliable index of poor intelligence, while if frequent, they indicate not only poor intelligence, but also the lack of moral qualities, such as self-control and carefulness. Incidentally, Meumann points out that, in theory, we should distinguish carefully between natural ability and actual ability as shown in school performance; in practise, nevertheless, these tend to coincide.

(15) *Memory of defectives.* Galton (18) applied Jacobs' tests to imbeciles, and found that most *S*'s of this type failed to repeat more than 4 digits, while several imbeciles who had remarkable memories for dates or for passages in books showed complete failure (span not over 3) in memory for digits. Johnson (23) computes the average span for feeble-minded (selected *S*'s of the so-called 'school-case' group) at 5.3, or approximately 1.3 digits less than the normal span of an 8-year child. The distribution of efficiency, as he found it, is shown in Table 79. Johnson comments upon the fact that the difference between the memory span

TABLE 79

Memory Span for Digits in the Feeble-Minded (Johnson)

Number of digits	3	4	5	6	7	8
Repeated correctly by	70	66	51	27	14	4

NOTE—The larger groups include the smaller ones at their right.

of the feeble-minded and of normal children seems to be of a smaller order than the general difference in intellectual ability of the two groups.

Miss Norsworthy compared normal and feeble-minded children with respect to memory for related and for unrelated words. Her standards for normal children have already been reported (Table 75): the relation of feeble-minded to normal efficiency is shown in Table 80. The figures are to be interpreted simply: five per cent of the feeble-minded do as well with the related-word test as do 50 per cent of normal children, etc.

Smedley states "that the boys of the John Worthy School [incorrigibles, defectives, truants, etc.] are lower in memory power

than are the pupils of the other schools, and this disparity increases with age" (43, p. 59).

Smith's tests with epileptics (46) show that, in the auditory letter-span test, they are generally inferior to normal *S*'s, and in particular, that they make nearly three times as many errors of insertion.

TABLE 80

Comparative Memory Capacity of Normal and Feeble-Minded Children (Norsworthy)

	ABOVE MEDIAN	ABOVE —1 P. E.	ABOVE —2 P. E.
Normal (both tests)	50	75	91
Feeble-minded, in related words	5	19	30
Feeble-minded, in unrelated words	6	18	27

(16) *Other correlations.* Krueger and Spearman found no correlation between memory for digits (serial visual exposure) and either ability to add, to discriminate pitch, or to discriminate dual cutaneous impressions.

Memory for digits and memory for letters were correlated to a high degree in Miss Sharp's *S*'s, while memory for short sentences correlated best with memory for letters.

Smedley studied the relation of memory for digits and ability to spell, and concluded that "while, on the whole, the good spellers have decidedly better memory power than the bad spellers, yet there are individuals among the poor spellers who are superior in memory power, and individuals among the best spellers whose memory power is scarcely up to the average of their age. While this native power of sense memory plays an important rôle, it is by no means the only factor in learning to spell" (43, p. 61).

TABLE 81

Recall of Different Members of a 7-Term Series (Binet and Henri)

Place in series	1	2	3	4	5	6	7
Times recalled correctly	143	139	115	111	122	117	140

(17) *Miscellaneous observations.* (*a*) Reproduction in correct order is more difficult than mere reproduction; reproduction is more difficult than recognition.

(*b*) The first and the last terms of a series are more liable to be recalled than are the middle terms (Table 81).

(*c*) In word tests, certain terms are often found to have a special reproducibility, evidently by attracting special attention in some way. Thus, Binet and Henri found that the word *pupitre* (desk), though in the middle of a series, and hence unfavorably placed, was recalled in an usually large number of cases.

(*d*) Errors of omission are more common than errors of insertion or errors of substitution—in word tests with school children, 4 times more frequent (Binet and Henri).

(*e*) Wissler calls attention to the perseverative tendency mentioned by Meumann and others: this is evinced by the introduction, in the recall of a given series, of impressions that had been used in an earlier series. Wissler found this type of error especially common in college seniors and mature *S*'s when trying the digit test. Meumann, it will be remembered, considered perseveration in the word test as an index of poor intelligence—when the *S*'s knew that no series was like a previous one.

REFERENCES

(1) I. M. Bentley, The memory image and its qualitative fidelity, in A. J. P., 11: 1899, 1–48.

(2) J. A. Bergström, Experiments on physiological memory by means of the interference of associations, in A. J. P., 5: 1893, 356–369. See also An experimental study of some of the conditions of mental activity, *ibid.*, 6: 1894, 267–273, and Relation of the interference and the practise effect of an association, *ibid.*, 6: 1894, 433–442.

(3) J. A. Bergström, Effect of changes in the time variables in memorizing, together with some discussion of the technique of memory experimentation, in A. J. P., 18: 1907, 206–238.

(4) A. Bernstein and T. Bogdanoff, Experimente über das Verhalten der Merkfähigkeit bei Schulkindern, in Beiträge z. Psych. d. Aussage, 2: 1905, 115–131.

(5) A. Binet, Attention et adaptation, in A. P., 6: 1899 (1900), 248–404.

(6) A. Binet, Introduction à la psychologie expérimentale, Paris, 1894. Especially ch. V.

(7) A. Binet and V. Henri, La mémoire des mots, in A. P., 1: 1894 (1895), 1–23.

(8) A. Binet and V. Henri, La psychologie individuelle, in A. P., 2: 1895 (1896), 411–465, especially 436–443.

(9) T. L. Bolton, The growth of memory in school children, in A. J. P., 4: 1892, 362–380.

(10) B. Bourdon, Influence de l'âge sur la mémoire immédiate, in Revue philos., 38: 1894, 148–167.

(11) W. H. Burnham, Memory historically and experimentally considered, in A. J. P., 2: 1888–89, 39–90, 225–270, 431–464, 568–622.

(12) C. Burt, Experimental tests of general intelligence, in B. J. P., 3: 1909, 94–177.

(13) Mary W. Calkins, A study of immediate and of delayed recall of the concrete and of the verbal, in P. R., 5: 1898, 451–6.

(14) J. Cohn, Experimentelle Untersuchungen über das Zusammenwirken des akustisch-motorischen u. des visuellen Gedächtnisses, in Z. P., 15: 1897, 161–183.

(15) H. Ebbinghaus, (*a*) Ueber eine neue Methode zur Prüfung geistiger Fähigkeiten in ihrer Anwendung bei Schulkindern, in Z. P., 13: 1897, 401–457. (*b*) For review of the literature, see Grundzüge der Psychologie I, 2d ed. 1905, 633-707.

(16) E. Ebert and E. Meumann, Ueber einige Grundfragen der Psych. der Uebungsphänomene im Bereiche des Gedächtnisses, in A. G. P., 4: 1905, 1–232.

(17) P. Ephrussi. Experimentelle Beiträge zur Lehre vom Gedächtnis, Berlin, 1906. Pp. 191.

(18) F. Galton, Supplementary notes on 'prehension' in idiots, in Mind, 12: 1887, 79–82.

(19) Eleanor Gamble, A study in memorizing various materials by the reconstruction method, P. R. Mon. No. 43, Sept., 1909. Pp. 210. (Wellesley College Studies in Psychology, No. 1.)

(20) C. J. Hawkins, Experiments on memory types, in P. R., 4: 1897, 289–294.

(21) V. Henri, Éducation de la mémoire, in A. P., 8: 1901 (1902), 1–48.

(22) J. Jacobs, Experiments on 'prehension,' in Mind, 12: 1887, 75–9.

(23) G. E. Johnson, Contribution to the psychology and pedagogy of feeble-minded children, in Pd. S., 3: 1895, 245–301, especially 268–273.

(24) A. Jost, Die Associationsfestigkeit in ihrer Abhängigkeit von der Verteilung der Wiederholungen, in Z. P., 14: 1897, 436–472.

(25) F. Kemsies, Gedächtnisuntersuchungen an Schülern, in Z. P. P., 2: 1900, 21–30, 84–95.

(26) E. A. Kirkpatrick, An experimental study of memory, in P. R., 1: 1894, 602–9.

(27) E. Kraepelin, Der psychologische Versuch in der Psychiatrie, in P. A., 1: 1896, 1–91, especially pp. 73 ff. (See also G. Aschaffenburg, *ibid.*, 209–299.)

(28) F. Krueger and C. Spearman, Die Korrelation zwischen verschiedenen geistigen Leistungsfähigkeiten, in Z. P., 44: 1907, 50–114.

(29) J. Larguier des Bancels, Sur les méthodes de mémorisation, in A. P., 8: 1901 (1902), 185–204.

(30) M. Lobsien, Experimentelle Untersuchungen über die Gedächtniss-entwickelung bei Schulkindern, in Z. P., 27: 1901, 34–76.

(31) E. Meumann, Intelligenzprüfungen an Kindern der Volksschule, in E. P., 1: 1905, 35–101.

(32) W. McDougall, On a new method for the study of concurrent mental operations and of mental fatigue, in B. J. P., 1: 1905, 435–445, especially 436 f.

(33) G. Müller and A. Pilzecker, Experimentelle Beiträge zur Lehre vom Gedächtniss, in Z. P., Ergänzungsband, 1: 1900, 1–288.

(34) G. Müller and F. Schumann, Experimentelle Beiträge zur Untersuchung des Gedächtnisses, in Z. P., 6: 1894, 81–190, 257–339.

(35) A. Netschajeff, Experimentelle Untersuchungen über die Gedächtnissentwickelung bei Schulkindern, in Z. P., 24: 1900, 321–351.

(36) Naomi Norsworthy, The psychology of mentally deficient children (Columbia University thesis), New York, 1906. Pp. 111.

(37) M. Offner, Das Gedächtnis, Berlin, 1909. Pp. 238.

(38) A. Pohlmann, Experimentelle Beiträge zur Lehre vom Gedächtniss, Berlin, 1906. Pp. 191. (For full review, see Z. P., 44: 1907, 134–140.)

(39) F. Reuther, Beiträge zur Gedächtnisforschung, in Psych. Studien, 1: 1906, 4–101.

(40) C. Ritter, Ermüdungsmessungen, in Z. P., 24: 1900, 401–444.

(41) M.-C. Schuyten, Sur les méthodes de mensuration de la fatigue des écoliers, in Ar. P., 4: 1904, 113–128.

(42) Stella E. Sharp, Individual psychology: a study in psychological method, in A. J. P., 10: 1899, 329–391.

(43) F. Smedley, Report dept. child-study and pedagogic investigation, (Chicago Public Schools), No. 3, 1900–1901, also in U. S., 1902, i., 1115–1138.

(44) W. G. Smith, The relation of attention to memory, in Mind, n. s. 4: 1895, 47–73.

(45) W. G. Smith, The place of repetition in memory, in P. R., 3: 1896, 21–31.

(46) W. G. Smith, A comparison of some mental and physical tests in their application to epileptic and to normal subjects, in B. J. P., 1: 1905, 240–260.

(47) E. B. Titchener, Experimental psychology, Vol. I., New York, 1901.

(48) J. J. van Biervliet, L'éducation de la mémoire à l'école, in Revue philos., 57: 1904, 569–586.

(49) R. Wessely, Zur Frage des Auswendiglernens, in Neue Jahrbücher f. Pädagogik, 8: 1905, 297–309, 373–386.

(50) W. H. Winch, Immediate memory in school children, in B. J. P., 1: 1904, 127–134, and 2: 1906, 52–57.

(51) W. H. Winch, The transfer of improvement in memory in school children, in B. J. P., 2: 1908, 284–293.

(52) C. Wissler, The correlation of mental and physical tests, in P. R. M. S., 3: 1901, No. 6. Pp. 62.

TEST 39

Memory for ideas: 'Logical' memory.—This test differs from the preceding tests of memory in two respects; in the first place, connected, meaningful material is used instead of a series of disparate impressions; in the second place, the reproduction that is demanded is primarily a reproduction of ideas, not an exact, verbatim reproduction of the original presentation. In other words, this test, to use popular phraseology, measures 'logical,' instead of 'rote' memory.

The purposes of the test are similar to those of other memory tests, viz:, to determine individual differences in memory efficiency, as related to sex, age, training, native ability, etc. As in those tests, too, the effect of different methods of presenting the material, or of different forms of material may be studied, and immediate may be compared with deferred reproduction. The results of the test may also be correlated with the results of other tests, particularly with the tests of rote memory just described.

While, in principle, the attitude taken by *S* toward the test of memory for ideas is distinctly different from that taken toward the test of memory for discrete impressions, yet, in practise, it is not always possible to differentiate these attitudes in the tests as actually administered. Thus, Binet and Henri, and after them, Miss Sharp, conducted tests of "memory for sentences." In these tests, the sentences ranged from short to long, and from easy to difficult. A short, easy sentence, *e.g.*, a sentence of 11 words, is almost invariably interpreted by *S* as a straightforward test of verbal memory, and the reproduction is at bottom a recall in verbal (mainly auditory verbal) terms. On the other hand, a long, difficult sentence, *e.g.*, a sentence of 86 words, when heard or read but once, must be reproduced in substance, not verbatim, and the recall, for most *S*'s at least, is a recall by meaning, a reproduction of the 'gist' of the material presented.

It is evidently better to keep separate these two different forms of memory test, with their two correspondingly different attitudes. The material of the present tests is, accordingly, sufficiently lengthy to preclude verbatim recall. Memory for sentences of progressive length may be studied by use of the sentences incorporated in the Binet-Simon tests of Chapter XIII.

In addition to the work of Binet and Henri and of Sharp, examples of the use of the logical-memory test may be found in Shaw's

study of memory in school children, in Wissler's records of the tests of Columbia freshmen, and in Terman's study of bright and dull boys. The most elaborate investigation of "memory for connected trains of thought" is, however, that of Henderson, who administered a series of tests to over 200 S's, ranging from 10-year old 5th-grade children to adult students in the university. Henderson's work forms the basis of the tests which are here prescribed, with some modifications suggested by the use of the test by the author for several years as a class exercise.

MATERIALS. — Watch. Three printed forms—The Marble Statue, Cicero, and The Dutch Homestead.

The first of these is taken from the appendix of Shaw's article, and was apparently used by him for subsidiary tests. The second and third are Nos. 2 and 4 of the five texts used by Henderson. If E wishes to extend the test by using more difficult material, he may employ Henderson's No. 5—a selection entitled "The Stages in the Development of Human Theory," from Comte's *Positive Philosophy*. If the Marble Statue proves too difficult or uninteresting for very young S's, E may read to them the following (from Clyde and Wallace, *Through the Year*, Book 2, Silver, Burdett, and Company).

HOW MR. LINCOLN HELPED THE PIG.

"One day Mr. Lincoln was out riding. As he passed along the road, he saw a pig sinking into a mud-hole. Poor piggy would climb part way up the slippery bank, then down he would fall again.

'I suppose I should get down and help that pig,' thought Mr. Lincoln. 'But I have on my new suit, and it will be quite spoiled if I do so. I think I'll let him get out the best way he can.'

He rode on. When nearly two miles away, he turned and came back. Not minding the new clothes, he stooped, and taking piggy in his arms, he dragged him out of the mud.

The new suit was quite spoiled, but Mr. Lincoln said he had taken a pain out of his mind."

METHOD.—Provide S with paper and pencil. Explain the nature of the test, as follows: "I am going to read you something to see how well you can remember it afterward. You must pay careful attention, as I shall read it but once. As soon as I have finished, take your pencil and write as much of the story as you can

remember. If you can remember it in just the words you heard, use those words, but if you can't do that, tell in your own words, as well as you can, what it was that I read to you."

Read the passage, including the title, with most careful enunciation, and with proper attention to expression. The rate of reading should be somewhat slower than in ordinary reading—say a full minute for the Cicero text. Allow *S* ample time for writing, then ask him to underline each word in his reproduction that he feels sure is exactly the same as the original passage.[1]

VARIATIONS OF METHOD.—(1) Supply *S* with the printed text. Inform him that he is to have 3 min. to read the passage. Assure him that this time is ample to read it over carefully several times. Direct him to read the passage straight through twice, and then to commit it to memory as he wishes.

(2) Defer the reproduction to any desired time after the reading, *e.g.*, 10 min., 24 hours, 1 week, 4 weeks. Or require an immediate reproduction, followed later, at one or more of the intervals just suggested, by a second or by a third reproduction.[2] Conduct these deferred trials in the same manner, as far as directions to underline, etc., are concerned, as in the first trial.

TREATMENT OF DATA.— The simplest plan for scoring the data of this test is that used by Terman and by Wissler, who merely graded the papers on a scale of 5 (or of 10) for a perfect reproduction—perfect in the sense of a reproduction of all the ideas of the original text, whether in terms identical with, or merely equivalent to the original.

For ordinary purposes, the author has found it serviceable to

[1] This test lends itself easily to the group method. The usual precautions should be taken to avoid disturbance and communication. *E* may save himself much labor by asking each *S* to count the total number of words he has written, then the total number of words he has underlined. With mature *S*'s, *E* may also reread very slowly the original text, and let each *S* check up the total number of ideas correctly reproduced, *i.e.*, represented, whether verbatim or by equivalent phrases, in his reproduction. The division of each text into its constituent 'ideas' is indicated below.

[2] It is better, on the whole, to give no intimation of the intent to demand a second reproduction. Some *S*'s may compare notes after the first reproduction, but if the subsequent trial is announced beforehand, coupled, as it ought to be, with the request not to think of the test in the interim, the request is more apt to work as a counter-suggestion, so that many *S*'s will test their recall of the passage, and otherwise furbish up the memories during the interval.

score the papers for the following points: (1) number of words written, (2) number of words underlined, (3) percentage of underlined words that are correctly underlined, (4) number of ideas ('details' in Henderson's terminology) that have been reproduced, whether exactly or in equivalent phrases. To these may be added, if desired, (5) number of ideas wrongly inserted. If but a single score is to be made, the fourth is obviously the one to be used, since the task assigned to *S* is to give as many as possible of the ideas of the text.[1]

The second and subsequent reproductions are scored in the same manner as the first. Retention is then measured, following Shaw and Henderson, by computing the percentage of loss between these and the first reproduction. Occasional cases of improvement in the later reproductions are rated as a zero loss.

To ascertain the 'idea-score,' *S*'s reproduction must be compared, step by step, with the standard divisions of the original text into ideas.[2]

The Marble Statue

(166 words, 67 ideas)

A young | man | worked | years | to carve | a white | marble | statue | of a beautiful | girl. | She grew prettier | day by day. | He began to love the statue | so well that | one day | he said to it: | "I would give | everything | in the world | if you would be alive | and be my wife." | Just then | the clock struck | twelve, | and the cold | stone began to grow warm, | the cheeks red, | the hair brown, | the lips to move. | She stepped down, | and he had his wish. | They lived happily | together | for years, | and three | beautiful | children were born. | One day | he was very tired, | and grew | so angry, | without cause, | that he struck her. | She wept, | kissed | each child | and her husband, | stepped back | upon the pedestal, | and slowly | grew cold, | pale | and stiff, | closed her eyes, |

[1] Consult Sharp or Henderson for more elaborate methods of treating data, particularly for devices for qualitative analysis.

[2] The scoring for ideas for these three passages is taken, with a few minor changes, from Shaw and from Henderson. For a division of the second and third texts into topics and sub-topics as well as into ideas, the reader may consult Henderson (2, pp. 29–30).

and when the clock | struck | midnight, | she was a statue | of pure | white | marble | as she had been | years before, | and could not hear | the sobs | of her husband | and children.

Cicero

(125 words, 65 ideas)

"Cicero, | the greatest | of the Roman | orators, | was born | at Arpinum, | an obscure | country | town.| His family | was of the middle class | only, | and without wealth, | yet he rose | rapidly | through the ranks | of Roman | official service | until at the age | of forty-six | he became | consul.| In oratory | he is | by universal consent | placed side by side | with Demosthenes, | or at least | close after him. | He surpassed | the great | Attic | orator | in brilliancy | and variety, | but lacked | his moral | earnestness | and consequent | impressiveness.| He could be | humorous,| sarcastic, | pathetic, | ironical, | satirical, | and when he was malignant | his mouth was | most | foul | and his bite | most | venomous. | His delivery | was impassioned | and fiery, | his voice | strong, | full, | and sweet, | his figure | tall, | graceful, | and impressive."

The Dutch Homestead

(180 words, 91 ideas)

"It was | one | of those spacious | farm- | houses, | with high- | ridged, | but lowly | sloping | roofs, | built | in the style | handed down from | the first | Dutch | settlers, | the low | projecting | eaves | forming a piazza | along the front | capable | of being closed up | in bad weather. | Under this | were hung | flails, | harness, | various | utensils | of husbandry, | and nets | for fishing | in the neighboring | river. | Benches | were built | along the side | for summer use; | and a great | spinning wheel | at one end, | and a churn | at the other, | showed | the various | uses | to which this important | porch | might be devoted. | From this piazza | one might enter | the hall, | which formed | the center | of the mansion | and the usual | place of residence. | Here | rows | of resplendent | pewter | ranged | on a long | dresser | dazzled | his eyes. | In one corner | stood a huge | bag | of wool,| ready | to be spun; | in another |

a quantity | of linsey-woolsey, | just | from the loom; | ears | of Indian | corn | and strings | of dried | apples | and peaches | hung | in gay | festoons | along the walls, | mingled | with the gaud | of red | peppers."

RESULTS.—(1) *Individual differences* in memory for ideas are unexpectedly large, even within a group of *S*'s of apparently similar attainments (Table 82).

(2) *Sex differences* in this test, as in the rote-memory test, are in favor of girls. The difference is indicated clearly in the author's

TABLE 82

Old Homestead Test. Words Written and Underlined (Whipple)

	FIRST TRIAL, NO INTERVAL		SECOND TRIAL, 24 HOURS LATER	
	Total Words	Words Underlined	Total Words	Words Underlined
Average, 9 men	80.4	48.0	83.0	38.4
Average, 22 women	95.5	38.8	99.6	34.1
Maximal records	127.0	102.0	138.0	66.0
Minimal records	45.0	4.0	52.0	9.0

data for college students (Table 82), and similar differences are reported by Shaw, who found the growth of memory for ideas to be faster in girls than in boys, and the average performance of girls to be some 4 per cent better than that of boys. Wissler's records for Columbia freshmen show an average of 44.5 per cent, P. E. 11.1, for men, and 48.2 per cent, P. E., 13.2, for women.

(3) *Age.* Binet and Henri state simply that, in the test with long sentences, the number of words retained is related to age. The more elaborate studies of Shaw and of Henderson are somewhat difficult to interpret. It appears evident, however, that a distinction must be made between efficiency in the first reproduction and efficiency in subsequent reproductions. If the first be termed learning capacity, and the second retentive capacity, and if the latter be measured in terms of the proportion of the first reproduction that is retained in the second (or later) reproduction, then adults may be shown to surpass children in learning capacity, but not in retentive capacity.

Thus, in Shaw's rather difficult 324-word story, the learning capacity of boys increased, from the 3d to the 9th grade, from 17 to 42 per cent, that of girls from 18 to 43 per cent. High-school boys averaged only 40 per cent, high-school girls about 47 per cent. Shaw's university students did no better, while Henderson's summer session students were inferior to his 15 and 16-year old school children. In short, then, logical, like rote memory, appears, when measured by the first reproduction, to be at its best near puberty.

Turning to the later reproductions, Shaw and Henderson (Table 83) agree that younger *S*'s have as good retentive capacity as do adults.

TABLE 83

Average Percentage of Loss in Third Reproduction (After 4 Weeks) (Henderson)

AGE	ADULTS	16	15	14	13	12	11	10
Percentage of loss....	14	8	13	15	14	12	10	10

(4) *Time-interval.* The insertion of a time-interval between presentation and reproduction has much less effect upon memory for ideas than upon memory for discrete impressions. Table 82 shows that, if a second reproduction is called for one day after the first, the average *S* actually writes more words. The words in the later reproduction are, however, less exact copies of the original text, and there is a tendency to insert extraneous material, so that fewer words are underlined, and there is a slight net reduction in the number of ideas reproduced. In the author's tests, this reduction was but 3 per cent at the end of one week. Table 83 shows that an interval of 4 weeks produces a loss of but 8 to 15 per cent.

It is a matter of special interest to note that the relative standing of *S*'s remains practically identical in tests conducted with immediate, and with deferred reproduction. It follows that, so far as this test goes at least, the popular notion "easy come, easy go" is not borne out by experimental evidence. Henderson found that this correlation between learning capacity and retentive capacity was brought out better in scoring for ideas than in scoring for words.

(5) *Dependence on method of presentation.* When a single hearing is compared with reading done by *S* (3 min.), the former is found to be nearly as good as the latter for immediate reproduction, but the latter to be much more effective than the former for deferred reproduction.

(6) *Dependence on practise.* There is little doubt that practise will improve memory for ideas, as it will improve nearly every form of psychophysical activity. Special training thus accounts, in all probability, for the high scores (52 as over against 40 to 47 per cent) reached by the pupils of Miss Aiken's school[1] in comparison with the work of Worcester high-school children.

(7) *Dependence on length of text.* The number of words reproduced after one hearing increases, though not in direct proportion, with the length of the passage heard (Binet and Henri).

(8) *Dependence on portion of text.* If the original passage be divided into 3ds or 4ths (or even, if long, into 8ths), it will be found that, on the average, the reproduction of any one of these portions is inferior to the one that precedes it and superior to the one that follows it. Thus, Shaw's story, on division into 4 parts, was found to be reproduced in the amounts 52, 34, 31, and 28 per cent, respectively.

(9) *Uniformity.* If several texts are used with the same group of *S*'s, their standing in the several trials shows a high degree of correlation. The test may, therefore, be regarded as having a good degree of reliability.

(10) *Relation with school standing.* In 86 cases, Wissler found a correlation of 0.19 with class standing, of 0.11 with standing in mathematics, and of 0.22 with standing in Latin. Henderson found but a slight correlation with class standing in the lower grades, but a closer correlation in the higher grades. He is of the opinion that, at least in the lower grades, the school marks put a premium upon industry and good conduct, rather than upon native ability, and thus obscure the existing correlation.

(11) *Other correlations.* Wissler found a correlation of 0.21 between logical memory and length of head, but no correlation between logical memory and rote memory, speed of naming colors, reaction time, or breadth of head.

[1] For an account of the special training given to Miss Aiken's pupils, see Test 25 and references thereto.

(12) *Qualitative aspects.* Inspection of the work of children and introspective examination by adults of their own mental processes reveal a number of interesting principles. In the first place, there is a process of selection: words or ideas that are logically or psychologically important are best retained. Or, as Henderson expresses it, there is, especially during a long time-interval, a process of condensation and generalization. The main ideas, the important topics, the brunt of the passage may remain fairly constant, but the minor details tend to be forgotten, and the original phrasing to become less and less clear.

When, then, the reproduction is demanded, most *S*'s first recall these main ideas or larger topics, and then develop the details, as best they may, from them. There is a strong tendency, in this filling out of the details, toward what Binet and Henri speak of as "verbal assimilation," *i.e.*, a tendency to express the ideas in one's own terms, rather than in those employed in the original passage. Thus, adults often use synonyms or other forms of substitution, while children replace the words of the original by words from their ordinary vocabulary (*e.g.*, *played* for *amused themselves*, *fire* for *conflagration*), and at the same time tend to simplify the syntax. In general, Binet and Henri found that the number of times that synonyms are used in the recall is, in short passages greater, and in long passages less than the number of ideas completely omitted.

Finally, the substitution of terms for those of the original tends, especially in younger *S*'s and with longer time-intervals, to become inexact; in other words, the sense of the original becomes more or less distorted. Thus, for instance, Binet and Henri discovered that, in all sentences containing more than 20 words, more than half of their *S's* had made some change in the meaning of the original. Of these alterations of sense, the most conspicuous are: (1) change of proper names or of numbers, (2) replacement of an object by an analogous object that might fit the sentence equally well, (3) insertion of details not inconsistent with the original, but still not in the original, and (4) alterations apparently due to emotional reaction, especially to exaggeration, *e.g.*, *a frightful snake* for *a snake*.

(13) *Miscellaneous points.* In the case of college students, from 50 to 90 per cent of the words underlined are actually correct. A certain type of *S* may be recognized, who is extremely cautious about underlining words, but who usually has these few nearly all correct.[1]

Binet and Henri estimate that memory for connected sentences is approximately 25 times as good as memory for discrete terms.

REFERENCES

(1) A. Binet and V. Henri, La mémoire des phrases, in A. P., 1: 1894 (1895), 24–59.

(2) E. N. Henderson, A study of memory for connected trains of thought. New York. Pp. 94. (Columbia University thesis.)

(3) Stella E. Sharp, Individual psychology: a study in psychological method, in A. J. P., 10: 1899, 329-391.

(4) J. C. Shaw, A test of memory in school children, in Pd. S., 4: 1896, 61–78.

(5) L. W. Terman, Genius and stupidity, in Pd.S., 13: 1906, 307–373.

(6) C. Wissler, The correlation of mental and physical tests, in P.R.M.S., 3: 1901, No. 16. Pp. 62.

---

[1] Cf. the type of cautious reporter mentioned in Test 32, p. 304.

CHAPTER X

Tests of Suggestibility

The term 'suggestion' has found different usages in psychology. Four different usages at least may be distinguished. (1) Suggestion is equivalent to association, *e.g.*, the idea 'horse' suggests the idea 'Black Beauty.' (2) Suggestion is the conveyance of an idea by hint, intimation, or insinuation, *e.g.*, the orator suggests an idea by an appropriate gesture. (3) Suggestion is a method of creating and controlling hypnosis. (4) Suggestion is a process of creating belief or affecting judgment, usually an erroneous belief or false judgment, in the normal consciousness.

The tests which follow all purport to measure susceptibility to suggestion in this last-named sense. In them, the experimenter seeks, by suitable arrangement of the test-material or of the instructions, to induce the subject to judge otherwise than he naturally would—to induce him, for example, to judge equal lines or equal weights to be unequal, or to perceive warmth when there is no warmth, etc. If the attempt is successful, the subject is said to have 'yielded,' or to have 'accepted' the suggestion; if unsuccessful, he is said to have 'resisted' the suggestion. The degree of his suggestibility is indicated by the quickness or frequency of his 'yields.'

Just as efficiency in observation, attention, memory, and the like has been shown to be specific, not general in character, so is it probable that suggestibility is specific, not general in character. For this reason, suggestibility must be tested by more than one method.

Many of the tests in other portions of this book, *e.g.*, Nos. 17 and 23, afford opportunity for noting the suggestibility of subjects. The serial graded tests of Binet and Simon (Chapter XIII) also contain directions for testing the suggestibility of young or of feeble-minded children.

TEST 40

Suggestion by the size-weight illusion.—Big things are ordinarily heavier than small things of the same kind. When we lift two weights of apparently the same material, but of different sizes, we more or less unconsciously put forth more energy or expect to meet with more resistance in lifting the larger. If, as in the case of the so-called 'suggestion-blocks,' the weights are really the same, we almost inevitably judge the larger weight to be the lighter; in other words, the visual appearance of the weight has given us a suggestion—or, as it turns out, rather, a disappointed suggestion,—of weight.[1]

This error of judgment is undoubtedly due to an association built up by long experience in handling and lifting various articles and objects.[2] One might, therefore, suppose that younger children, or less intelligent children, who would, presumably, have had less of this discriminative association of size and weight, would be less affected by the suggestion. For this reason, the size-weight test has been applied by several investigators to determine or to measure, at least relatively, the degree of suggestibility exhibited by school children under various conditions.

Apparatus.—Low table. Soft black cloth. Set of 'suggestion-blocks,' patterned after Gilbert, but modified by extending the comparison series in both directions.

This set consists of two standard blocks and 20 comparison blocks. Both standards weigh 55 grams; both are 28 mm. thick, but the larger is 82 and the smaller 22 mm. in diameter. The 20 comparison blocks are all 28 mm. thick and 35 mm. in diameter, but their weights range from 5 to 100 g. by 5 g. increments.[3] All are painted dead black.

---

[1] As Scripture remarks, the poor fellow who has been laughed at for centuries for saying that a pound of lead is heavier than a pound of feathers is perfectly right, so long as he speaks psychologically, and looks at the pillow and the bit of lead pipe. A concrete demonstration of the truth of this statement is afforded by several experiments reported by Wolfe.

[2] Some writers, however, *e.g.*, Flournoy, attribute the illusion to an inborn nervous connection. For a discussion of the psychological factors concerned in this experiment, particularly in its relation to the 'innervation-sense,' consult Flournoy, Müller and Schumann, Seashore, Bolton, Loomis, and van Biervliet.

[3] Gilbert's comparison blocks were but 14 in number, with a range from 15 to 80 g. This range proved inadequate for younger *S*'s.

Method.—Arrange the table at such a height that S's forearm will be parallel with the floor when lifting a weight. Spread the black cloth over the table: this cloth should be large enough to cover at least the portion of the table occupied by the weights, and thick enough to deaden the sounds incident to their replacement.

Arrange the twenty comparison blocks on the cloth, in the order of their weight from left to right, and in such a manner that any one of them may be reached by S without materially changing the angle of his arm. Place before S the larger standard block, and say: "Here is a block. I want you to find a block in this series of 20 blocks that seems to you just as heavy as this one. Lift it by picking it up edgewise with your thumb and finger, like this. [Illustrate.] Then try the first of these weights [at the left]. If that doesn't suit, try the next, then the third, and so on, till you find a block that seems equal to this one. Each time you must lift this block first, then the one you are trying in the series. Keep your eyes constantly directed at the weight you are lifting." When S has selected an equivalent weight, the same procedure is followed with the second, or smaller, standard block.

Our estimate of the absolute or relative weight of a body is conditioned by an unsuspectedly large number of factors, the analysis of which has been the occasion of a number of extended and carefully-executed researches.[1] While space forbids a discussion of these factors here, it should be impressed upon E that the conditions under which S lifts the blocks should be kept as uniform as possible. Particularly, since the apparent weight of a body depends in part upon the velocity and height to which it is lifted, it is important that S should pick up each block in the same manner, lift it at the same tempo and to the same height. Again, since the memory image for weight changes rapidly, S's judgment, in so far as it is based upon the image of the first weight, would be appreciably altered if the second weight were lifted at varying intervals after the first: the interval should, accordingly, be made as constant as possible, and fairly short, say not over 3 sec., and the arrangement of the weights must be such as to permit this procedure. Finally, in this test, since the suggestion hinges upon the visual perception of the block, E must be sure that S looks directly at each block as he lifts it.

---

[1] See particularly, Martin and Müller, and Müller and Schumann.

TREATMENT OF DATA.—Following Gilbert, Scripture, and Seashore, the force of suggestion produced by the difference in size of the two standard blocks may be indicated by the difference in weight, in grams, between the two comparison blocks that are selected by *S* as the equivalents of the two standards.

The force of the size-weight illusion has been expressed by Scripture, on the basis of the more elaborate suggestion-blocks used by Seashore, in the form of the following law:

$$i = \frac{k}{s+c} - d,$$

in which

i = the amount of the illusion,
s = the difference in size acting as a suggestion,
c = the diameter of the blocks of constant size,
d = the weight of the blocks of constant weight, and
k = a constant depending on the nature of the experiment (whether the blocks are directly observed, whether *S* knows the nature of the illusion, etc.).[1]

RESULTS.—(1) The effect of *age* upon suggestibility by the size-weight illusion is judged by Dresslar, who employed another and less satisfactory method, to be indifferent. Gilbert, however, found, as Table 84 indicates, that the illusion is strong at 6, increases gradually till 9, and thence declines with age.

(2) The *relation of sex* to suggestion by the size-weight illusion has been differently stated by different investigators. Dresslar, for example, concludes that boys are more suggestible than girls. Wolfe, on the contrary, states that "men are less prone than women to illusions of weight," and that, in comparing wooden with lead weights, "the women overestimate the lead nearly twice as much as the men." Gilbert and Seashore find females more suggestible than males, but in nothing like the degree stated by Wolfe. Thus, inspection of Table 84 shows that, according to

[1] For the data from which this law is derived, see Scripture (15, p. 276 f.), also Seashore (11, pp. 3–14). For a striking demonstration of the force of the illusion, reference may be made to Wolfe's statements that "about one woman in 7 finds 1 g. of lead equal in weight to 60 g. of inflated paper bag," and not "one woman in 7 will find a gram of inflated paper bag half as heavy as a gram of lead" (p. 460).

Gilbert's method, girls are, after the age of 9, on the average, more influenced by the illusion than are boys. Seashore (12) tested 17 women and 28 men with two test-weights quite different in size, and found, similarly, that on the average the women showed the stronger illusion.

TABLE 84

Force of Suggestion (Gilbert)

AGE	6	7	8	9	10	11	12	13	14	15	16	17
NB.......	45	50	46	47	49	43	54	45	47	49	47	43
NG	47	45	46	47	42	48	49	58	53	51	39	41
F	42.0	45.0	47.5	50.0	43.5	40.0	40.5	38.0	34.5	35.0	34.5	27.0
P	36	37	27	36	23	22	15	8	7	12	6	5
MV	17.0	15.5	13.5	10.5	12.5	11.5	9.0	9.0	9.5	10.5	10.0	12.0
FB	43.5	43.5	45.0	50.0	40.0	38.5	38.0	37.0	31.0	33.0	32.0	25.0
FG	42.0	43.5	49.5	49.5	44.0	40.0	41.0	38.0	33.5	38.0	38.5	31.0

NB = number of boys
NG = number of girls
F = force of suggestion, in grams, for both sexes (median values)
FB = force of suggestion, in grams for boys (median values)
FG = force of suggestion, in grams, for girls (median values)
P = per cent. of cases in which *F* exceeded 65 g., the limit used
MV = statistical mean variation

Gilbert's explanation is given in the following terms: "At 6 he has not yet learned to compare. As he learns gradually to judge a thing from more aspects than one, or in other words, learns to interpret one sense by another, the force of suggestion given by the eye to the muscle increases until at 9 he has come to the age of experience enough to see that things are not always what they seem. Consequently at this age he begins to correct misleading influences bearing upon him."

(3) *Practise*, even if regular and persistent, does not dispel the illusion. If *S* be told the nature of the illusion, it still persists, though its intensity is thereby somewhat reduced (Seashore).

(4) The *relation of intelligence* to suggestibility has not been treated as carefully as the problem warrants. Gilbert made no correlations with intelligence. Dresslar concluded that bright children exhibit a stronger illusion, but Seashore (11) contends

that Dresslar's method (arrangement in serial order) did not afford a real measure of the strength of the illusion.

(5) If the method of procedure be modified, the strength of the illusion will be altered.

The more important of the relations thus revealed are the following:[1]

(*a*) "The illusion of weight dependent on size is greatest when size is estimated mainly by muscle-sense, and the weights have not previously been seen."

(*b*) "The illusion is more fluctuating and on the whole not quite so strong when size is estimated by the area of pressure in the flat palm, including a memory of the third dimension."

(*c*) "In these variations, the illusion is weakest when size is estimated by direct sight."

(*d*) "When size is estimated by the combined effect of all the spatial senses, the illusion is weaker than when depending on muscle-sense or touch, and stronger than when dependent on sight alone."

(*e*) The illusion is weaker when the blocks are viewed in indirect vision, and still weaker when judged by visual memory.

(*f*) A knowledge, or supposed knowledge, of the material of which weights are made may affect the estimate of their weight.

(*g*) The illusion does not necessarily vary directly with the volume of the compared weights, but depends in part upon the manner in which the difference in volume is brought about.

(*h*) The illusion obtains among the blind, where it follows the same general law as for the seeing, though it is not so strong, either for lifted or merely 'touched' weights, as for the seeing under the same conditions (Rice).

NOTES.—The outcome of any test of weight-comparison is somewhat affected by the tendency felt by all *S*'s, though differing in degree between different individuals and in the same individual at different times, to overestimate the second of two lifted weights.

If blocks of different material, *e.g.*, cork and lead, or wood and iron, be constructed in such a manner as to have the same dimensions and the same weight, the knowledge of the actual differences in the weight of the two materials produces an illusion similar to the size-weight illusion. Seashore (12) tested school children with this material-weight illusion, and found that the overestimation of the metal blocks amounted to from 7 to 11 grams (or from 13 to 20 per cent of their actual weight, 55 g.). For this illusion, it

[1] See, especially, Seashore (11).

is of interest to note, there was found virtually no variation with age, sex, or intellectual ability.

REFERENCES

(1) F. E. Bolton, A contribution to the study of illusions, etc., in A. J. P., 9: 1898, 167–182, especially 167–178.

(2) A. Charpentier, Analyse expérimentale de quelques éléments de la sensation de poids, in Archives de physiologie normale et pathologique, 5th ser., 3: 1891, 122–135, especially 126 ff.

(3) F. B. Dresslar, Studies in the psychology of touch, in A. J. P., 6:1894, 313–368, especially 343–360.

(4) Th. Flournoy, De l'influence de la perception visuelle des corps sur leur poids apparent, in A. P., 1: 1894 (1895), 198–208.

(5) J. A. Gilbert, Researches on the mental and physical development of school children, in Y. S., 2: 1894, 40–100, especially 43–5, and 59–63.

(6) H. N. Loomis, Reactions to equal weights of unequal size, in Y. S., n. s. 1: No. 2, June, 1907, 334–348. (Same as P. R. M. S., 8: No. 3, whole No. 34.)

(7) L. Martin and G. E. Müller, Zur Analyse der Unterschiedsempfindlichkeit, Leipzig, 1899. Pp. 233.

(8) G. E. Müller and F. Schumann, Ueber die psychologischen Grundlagen der Vergleichung gehobener Gewichte, in Archiv f. d. ges. Physiol., 45: 1889, 37–112.

(9) J. Philippe and J. Clavière, Sur une illusion musculaire, in Revue philos., 40: 1895, 672–682.

(10) J. F. Rice, The size-weight illusion among the blind, in Y. S., 5: 1897, 81–7.

(11) C. E. Seashore, Measurements of illusions and hallucinations in normal life, in Y. S., 3: 1895, 1–67, especially 1–29.

(12) C. E. Seashore, The material-weight illusion, in I. S., 2: 1899, 36–46.

(13) E. W. Scripture, Remarks on Dr. Gilbert's article, in Y. S., 2: 1894, 101–4.

(14) E. W. Scripture, The law of size-weight suggestion, in Science, n. s. 5: Feb. 5, 1896, 227.

(15) E. W. Scripture, The New Psychology, London, 1897. See ch. xix.

(16) J. van Biervliet, La mesure des illusions de poids, in A. P., 2: 1895, 79–86.

(17) H. K. Wolfe, Some effects of size on judgments of weight, in P. R., 5: 1898, 25–54.

TEST 41

Suggestion by progressive weights.—This test, like that which follows it, is one of several devised by Binet for the purpose of

securing a quantitative measure of the degree of suggestibility of children or adults when the suggestion is 'depersonalized,' in the sense that it is derived by *S* himself from the objective conditions of the experiment, rather than from the attitude, tone, instructions, or personality of *E*. The principle embodied in this test is, in other words, the arousal, by auto-suggestion, of a "directive idea," or the rapid development of an attitude of expectation. Suggestibility is measured, at least approximately, by the ease with which this suggestion, or habit of judgment, is aroused and by the persistence that it displays under conditions which tend gradually to counteract it.

MATERIALS.—A set of 15 weights, of identical size and appearance, numbered conspicuously from 1 to 15. The first four weigh 20, 40, 60, and 80 grams, respectively; the remaining 11 weigh 100 grams each. Table of such a height that *S* can stand in front of it and lift the weights readily. A thick gray or black cloth.

PRELIMINARIES.—Spread the cloth over the table. Place the 15 weights on it in a line, with the lightest on the left and the 11 heaviest on the right, and with about 2 cm. between each weight. No. 1 is then at the left, No. 15 at the right, of the row.

METHOD.—Give *S* the following instructions: "Here is a series of weights, 15 of them. I want you to lift them, one after the other, like this. [Illustrate by taking a weight between thumb and finger and lifting some 10 cm. from the table.] As you lift each weight, I want you to tell me whether it is heavier, lighter, or the same as the one just before it. All you have to say is either 'heavier,' or 'lighter,' or 'the same.' Remember you are to compare each weight with the one you lifted just before. For instance, when you lift the 8th, you are to say whether it is heavier, lighter, or the same as the 7th. Here is the first weight, number one, at the left end of the row."

Watch *S* to see that he follows these instructions, particularly that he lifts the weights successively, without relifting earlier ones. Record his judgments verbatim; be careful, also, to note any secondary evidences that might throw light on his judgments, *e.g.*, attitudes or expressions of hesitancy, assurance, surprise, embarrassment, cautiousness, etc.

VARIATIONS OF METHOD.—(1) In the second method followed by

Binet, S is instructed to lift, in each trial, the preceding weight as well as the one that is being judged, *e.g.*, he lifts the 8th, then the 7th, then the 8th again: next the 9th, then the 8th, then the 9th again, etc. The lifting is all done, as before, with the one hand.

(2) In the third method followed by Binet, S is asked to estimate the first weight lifted. He usually gives too small an estimate. He is then told that its weight is 20 grams (about 0.7 ounce). The series is now compared, using either of the methods of lifting above described, according to S's preference,[1] but S is required to estimate or guess the heaviness of each weight, basing his judgment, of course, merely on the knowledge that the first weight is 20 grams.

Treatment of Data.—From the tabulated results, E may easily determine in how many cases the objective progression of the first 5 weights was correctly noted. For a measure of suggestibility, E must take the number of times 'heavier' is judged in the last 10 judgments (when 'same' is the correct judgment). This measure is admittedly somewhat crude, but it affords a fairly reliable index for determining the relative order of rank of a group of S's. Thus, an S that judges 'heavier' 10 times is unquestionably more suggestible than one who answers 'heavier' but 5 times, though not necessarily twice as suggestible.

If all three methods are employed, E may determine S's suggestibility by adding the number of false 'heavier' judgments in all three tests. In the third method, the quantitative estimate given by S for the 15th weight (or the maximal estimate for weights 6 to 15) might be taken, in comparison with his estimate of the 5th weight, as an index of suggestibility, but this method is not regarded by Binet as so reliable as the one already described.

Results.—(1) The general outcome of the test as conducted by the first, or standard, method is indicated in Table 85, which embodies the results obtained by Binet upon 24 elementary-school children, aged 8–10 years.

(2) It is evident that, in children of this age (8–10), not all judge correctly the actual objective increase in the *first five weights*. Since the differences are supraliminal, the exceptions must be

[1] It would, obviously, be better to prescribe either the one or the other method for all S's. The first method has the merit of taking less time, and it is the method that is for the most part naturally adopted by younger S's.

TABLE 85

The Progressive-Weight Suggestion. 24 Cases (Binet)

NO. OF WEIGHT	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Actual weight	20	40	60	80	100	100	100	100	100	100	100	100	100	100	100
Times estimated +		24	19	19	23	13	18	18	18	12	19	19	17	15	12
Times estimated −		0	1	1	0	9	1	4	1	7	2	2	3	3	5
Times estimated =		0	4	4	1	2	5	2	5	5	3	3	4	6	7

ascribed to faulty attention, though, possibly, the fact that the weights are of equal size may have clouded the direct perception of weight by lifting.

(3) In general, the suggestion is still working, though less powerfully, at the 15th trial: in other words, it has persisted, for most *S*'s, through the successive lifting of 10 equal weights.

(4) There is a marked drop in the judgment 'heavier' at *the 6th weight, i.e.,* at the first 'trick' weight—a drop which is, obviously, due to a 'disappointed suggestion,' analogous to that which conditions the size-weight illusion of the preceding test. In the present instance, *S* is, in most cases at least, prepared to find the 6th weight heavier than the 5th: he puts forth more effort; the weight rises with unexpected ease, and is, therefore, often judged 'lighter.' If, however, *S* is more influenced by his expectation of 'heavier' than by the unexpected lightness of the weight, he still judges 'heavier,' or he may, from the conflict of these two tendencies, judge 'equal.'

(5) *Practise* has very little effect upon the suggestibility of *S*'s: at least Binet found that, when 12 older children (16 years) repeated the test by the first method five times in immediate succession, there was no alteration in the average number of times that suggestion appeared (the average number of suggestions in the five trials was 5.1, 4.9, 5.4, 5.0 and 5.5, respectively).

(6) Tentative experiments indicate that *age* apparently has less effect upon suggestion by progressive weights than upon suggestion by progressive lines (see the following test): in trials by the first method, 12 children aged 16 years responded, on the average, with 5.1 suggestions, whereas 24 children aged 8–10 years, responded, on the average, with 6.75 suggestions.

(7) According to Binet, comparison of the results of this test with *other tests of suggestibility*, especially the line-test, indicates a fair degree of correlation, so that, while the sense-department under examination may in part determine the extent of suggestion, very suggestible *S*'s may be expected to prove noticeably suggestible in all tests. On the other hand, tests undertaken at the Educational Laboratory at Cornell University[1] do not confirm Binet's statement, and lead one to believe that Scott's conclusions (Test 44) are correct, when he asserts that there is no such thing as general suggestibility.

(8) Procedure by the *second method* (compulsory lifting of the antecedent weight) makes the real progression (1st five weights) more uniformly evident, but reduces the illusory progression.

(9) Procedure by the *third method* (estimates of each weight) produces less suggestion than the first, but more than the second, method. Inspection of the estimated weights (grams) shows (*a*) that *S*'s have a decided preference for the use of numbers terminating in 0 or 5, (*b*) that no one of 24 *S*'s overestimated the 5th weight (100 g.), but that they commonly greatly underestimated it (30 to 50 g.), and (*c*) that those *S*'s that showed the greater number of suggestions also gave, on the average, the largest quantitative estimations for the illusory increments: the correlation of suggestibility for these two methods of treatment was found by Okabe and Whipple to be 0.53.

REFERENCE

A Binet, La suggestibilité, Paris, 1900. Ch. iv. (pp. 261-208).

TEST 42

Suggestion by progressive lines.—The purpose and general plan of this test are the same as in the preceding test of suggestion by progressive weights, and the details are again derived from the work of Binet.

[1] These tests, which were conducted by T. Okabe, under the author's directions, included all the suggestibility tests of Binet, together with the warmth tests (No. 44). The results of their application to 29 *S*'s indicate almost total lack of correlation of suggestibility in the several tests. A full account will appear in the *Journal of Educational Psychology*.

APPARATUS.—A sheet of cross-section paper, ruled in millimeter squares. Kymograph drum, with kymograph or some form of supporting stand. Cardboard. Strip of white paper, 15 × 50 cm. Drawing utensils.

PRELIMINARIES.—Arrange the kymograph drum so that it may lie horizontally and be revolved freely by hand. It may conveniently be left in the kymograph with the driving 'step' loosened, or be placed in the smoking stand. On the strip of white paper, draw with a ruling pen 20 parallel, straight, black lines, 2 cm. apart and each 1 mm. wide. The lines must begin at varying distances from the left-hand margin: the first four are to be 12, 24, 36, and 48 mm. long, respectively; the remaining 16 are to be each 60 mm. long. Support the sheet of cardboard vertically in front of, and close to the kymograph drum, and cut a horizontal slit 1 × 12 cm. through the cardboard in such a position as to expose the ruled lines, one by one, as they are turned past the slit.[1]

METHOD.—Seat *S* 50 cm. from the screen and provide him with a sheet of cross-section paper. The instructions should take the following form. "I want to try a test to see how good your 'eye' is. I'll show you a line, say an inch or two long, and I want you to reproduce it right afterwards from memory. Some persons make bad mistakes; they may make a line 2 inches long when I show them one 3 inches long; others make one 4 or 5 inches long. Let's see how well you can do. I shall show the line to you through this slit. Take just one look at it, then make a mark on this paper [cross-section paper] just the distance from this edge [left-hand margin] that the line is long. When that is done, I shall show you the second line, then the third, and so on. Make the marks for the second on the line below the first, the third on the next line, and so on."[2]

[1] In default of the kymograph, the strip of ruled lines may be laid flat upon the table and exposed through a 1 × 12 cm. slit cut in the center of a sheet of cardboard 55 cm. square.

Or, the test-lines may be drawn as sections of radii upon a cardboard disc which is supported vertically just behind the screen and rotated to bring them into view successively.

[2] These directions should be followed with some care. In tests of suggestion, the slightest change in the setting of the test, or in the manner or content of the instructions, may materially affect *S*'s attitude toward the experiment. The object is to convey the idea of a straightforward test of accuracy of line-reproduction, and to avoid arousing any suspicion of snares or tricks.

E then turns the drum to bring the first, or shortest, line into view. As soon as *S* turns his attention to the recording of his estimate on the paper, the drum is moved forward slightly to conceal the line so that further comparison is impossible. As soon as *S* has placed his mark, then, and not before, the next line is exposed. This precaution serves to maintain the impression that a new, and hence probably a longer, line is exposed. Slow *S*'s may need to be hurried; too quick ones may need to be checked, so that the interval between successive exposures shall be approximately 7 sec. To keep *S*'s attention alive, *E* may accompany the exposures with non-suggestive remarks, *e.g.*, "Here is the second line." "Here is the third," etc.

If *S* has ceased to respond to the suggestion of progressive augmentation at the 20th exposure, the test ends at that point: if not, *E* should, without *S*'s knowledge, bring the drum back to the 5th line, and continue the exposures of the series of 60 mm. lines as before, until *S* does cease to respond to the suggestion.

E should note and record any significant features in *S*'s manner, *e.g.*, signs of embarrassment, hesitancy, automatic response, etc.

When the test is completed, and provided no further tests of suggestibility are to be undertaken at the time, *E* will find it advantageous to quiz *S* with regard to his attitude toward the test. This interrogation must be very tactfully conducted. *E* may, for example, begin by asking: "Are you entirely satisfied with what you have done"? If *S* answers in the affirmative, let *E* continue with such inquiries as: "Do you think you have made any mistakes"? "Did you make any lines too short or too long"? "At what moment did you notice that your lines were too long"? "Why didn't you make them shorter"? etc. If *S* confesses that he made some mistakes, let him take his record-sheet and make the changes that he thinks ought to be made to produce a correct record, using small circles for his corrections to avoid confusion with his first estimates.

Variations of Method.—*E* may, if desired, adopt the arrangement first used by Binet, according to which there are 12 successive stimulus-lines, all of which begin at the same distance from the left-hand margin, and which have the following lengths: 12, 24, 36, 48, 60, *60*, 72, *72*, 84, *84*, 96, *96* mm. It is evident that numbers

6, 8, 10, and 12 constitute four 'trap-lines,' since the arrangement suggests progressive augmentation, whereas each of these four lines is equal to that which immediately precedes it.

TREATMENT OF DATA.—(1) For a measure of suggestibility, *E* may take the number of lines out of the last 10 lines that are drawn longer than the 5th line was drawn.

(2) A coefficient of suggestibility may also be calculated, following Binet's method, by the formula

$$x : 100 = max.\ L : 5th\ L,$$

in which

x = the required coefficient,
$max.\ L$ = the length of the maximal line recorded by S,
$5th\ L$ = the length of the 5th line as recorded by S.

Absence of suggestibility is, then, indicated by a coefficient of 100: presence of suggestibility by a coefficient of over 100.

(3) When the variant method is used, the degree of suggestibility may be determined roughly in terms of the number of 'traps' in which S is 'caught,' or more exactly, by the formula

$$x : 100 = c : r,$$

in which

x = the required coefficient,
c = the average recorded increment of the four trap-lines,
r = the average recorded increment of the four lines immediately preceding the four trap-lines.

RESULTS.—(1) In his examination of pupils in the elementary schools, aged 8–10 years, Binet found that *the coefficient of suggestibility* ranged from 109 to 625. In 16 of 42 pupils, the coefficient was 200 or over, *i e.*, the maximal line was double or more than double the 5th line.[1]

With the variant form of test, Binet found the coefficient lying between 7.6 and 120. No one of 45 children avoided all four 'traps,' and 36 children avoided none of them. Occasionally, the trap-line, presumably on account of the contrast between the stimulus and the child's expectation, was actually recorded as shorter than the preceding line. *S*'s whose coefficient in this form of test is 100, *i.e.*, whose average increment for the trap lines is the same as for the objectively progressive lines, are termed 'automatic.'

[1] For detailed records of a number of individual cases, consult Binet, 124 ff.

(2) The point at which *maximal suggestibility* is registered is commonly between the 19th and the 25th line, but may lie anywhere between the 7th and the 36th (this being the limit tested by Binet).

(3) Inspection of the records of individual pupils shows that in some cases the force of suggestion was steady and persistent, while in others it reached a maximum, and then declined.

(4) *Extremely suggestible S*'s may make their 'estimate' of the line without even looking at it when exposed; their minds are so completely dominated by the suggestion of uniform augmentation that they do not trouble to observe the stimulus.

(5) The degree of suggestion induced by this test declines markedly with *age:* Binet found, for instance, that the coefficients of suggestibility, in the case of 12 pupils whose age averaged 16 years, ranged only from 103 to 146, and the author has not been able to produce appreciable suggestion in scattered tests of college students.

(6) In either form of test, the *1st line* is apt to be overestimated. The *5th line* is almost invariably underestimated. Generally speaking, this underestimation is less pronounced in those *S*'s that prove least suggestible.

(7) In many instances, the records bear witness to a struggle between the directive idea of progressive increments and the impressions which are actually received from the lines as they are exposed. Especially characteristic is the appearance of a number of estimates in which the directive idea is effective, followed by a sudden reduction in estimation, which is again followed by another series of progressive increments. In other words, the idea of progression is operative until a point is reached when the recorded length is manifestly too long. *S* makes, then, a more or less marked correction, but does not, curiously, relinquish the notion of progression, and this again becomes manifest.

(8) The corrections made by young *S*'s during the inquiry that follows the test cannot, of course, be taken as exact indications of the extent of the suggestion or of their consciousness of error. It will be found that many *S*'s are conscious that they have made the lines too long; some can also explain why they made them too long, but it is rare that any one gives a satisfactory explanation of why

he continued to make them too long, after he realized that he had been overestimating.

(9) *Correlations.* Tests of school children and of adults by Okabe and Whipple afforded the following correlations (footrule method): Suggestibility for progressive lines (number of 'yields') and suggestibility for progressive lines (maximal divided by the 5th line) 0.38. Correlation, by either treatment, with contradictory suggestion (Test 43) about 0.25, with directive suggestion (Test 43) about 0.20, with suggestion for warmth 0.17, with the size-weight illusion (Test 40) 0.10 by the first, and −0.14 by the second method of computing suggestion for progressive lines.

REFERENCE

A. Binet, La suggestibilité, Paris, 1900, pp. 83–160.

TEST 43

Suggestion of line-lengths by personal influence.—In the three preceding tests suggestion is produced by the objective conditions of the test: in everyday life, however, suggestion is often produced by personal influence, by authoritative statement or command, or merely by what Binet terms 'moral influence.' Two forms of line-test have been utilized by Binet to study this variety of personal suggestion: the first he terms 'contradictory suggestion,' the second 'directive suggestion' (*suggestion directrice*): in the former *E* makes certain statements that are intended to interrupt or modify a judgment that *S* has just made; in the latter, statements that are intended to control or influence a judgment that *S* is just about to make.

A. CONTRADICTORY SUGGESTION

MATERIALS.—Drawing utensils. A sheet of cardboard upon which are drawn in ink 24 parallel, straight, black lines, ranging in length from 12 to 104 mm., by increments of 4 mm. The lines all begin at the same distance from the left-hand margin, are 7

mm. apart, and are numbered in order of their length, from 1 to 24. Three rectangular pieces of cardboard, about 12 × 20 cm., on each of which is drawn a single straight line. These three stimulus-lines correspond to numbers 6, 12, and 18 of the 24 comparison-lines, and are, accordingly, 32, 56, and 80 mm. long, respectively.

METHOD.—Show *S* the card of comparison-lines, and explain their numbering. Replace this by the first stimulus-line (32 mm.), saying: "Look carefully at this line." After 4 sec., remove the stimulus-card, present the comparison-card, and say: "Tell me the number of the line that is just the length of the one I showed you." At the moment that *S* gives his judgment, *E* says: "Are you sure? Isn't it the —th"?—indicating always the next longer line. If S answers "No," *E* repeats the question in exactly the same form. If *S* still answers "No," the attempt to produce suggestion is suspended, and the case is recorded as one 'resistance.' The second and the third stimulus-lines are presented and the same procedure is followed in each case. If, in any of the trials, *S* answers "Yes," *E* then inquires: "Isn't it this one"?—indicating the next longer line, and this inquiry is carried on from line to line until *S* has twice resisted the suggestion, *i.e.*, has twice answered "No" to the same question.[1]

TREATMENT OF DATA.—Following Binet, *S*'s suggestibility may be rated in terms of the total number of 'advances' in lines that he makes, under inquiry, in all three trials. Thus, if he 'yields' two lines the first time, three the second, and none the third, his suggestibility is rated as 5.

RESULTS.—(1) Children tend to select for their first line one that is shorter than the stimulus-line.[2]

(2) Of 25 children, aged 8–10 years, Binet found 6 who resisted suggestion completely, 6 who 'yielded' once, 5 twice, 2 three times, 2 four times, and one each six, seven, and more than seven times.

[1] Once more it should be said that it is highly important to follow the same form of inquiry, to use the same tone, the same attitude, in every question for every *S*, since the suggestion which we seek to measure is conditioned by the character of the inquiries.

[2] *E* is almost always, therefore, in a position to demonstrate to *S*, if need be, after the test, that his suggestion would have been a sound one to follow.

(3) Preliminary experiments conducted by Binet and Henri upon 240 pupils, with some slight changes in method (particularly, giving an opportunity both for direct comparison and for selection by memory after a 12 sec. interval), yielded the results (2: p. 343) indicated in Table 86.

TABLE 86

Percentage of 'Yields' to Contradictory Suggestion (Binet and Henri)

AVERAGE AGE	MEMORY TEST	COMPARISON TEST	MEAN
7–9	89	74	81.5
9–11	80	73	76.5
11–13	54	48	51.0

Here it is evident that *E*'s suggestion is less effective when *S* can make direct comparison of the lines, and that suggestibility, under either direct comparison or comparison from memory, declines with *age*.

(4) *S*'s who have selected the correct line are less apt to change their designation under suggestion than are *S*'s who have selected the wrong line: thus Binet and Henri found that 56 per cent changed their selection when it was actually right, but 88 per cent when it was wrong. Moreover, of the latter, 81 per cent made the change in the proper direction.

B. DIRECTIVE SUGGESTION

APPARATUS.—As in Test 42, save that only the 60 mm. lines are used.

METHOD.—Seat *S* 50 cm. from the cardboard screen and provide him with a sheet of cross-section paper. Instruct him as follows: "I'm going to show you a number of lines. You will see them appear through this slit, one at a time. When I show you a line, take a good look at it; then make a mark on this paper at just the distance from this edge [left-hand] that the line is long. When that is done, I shall show you the second, then the third, and so on. You will make the mark for the length of the second line on the second line of your paper, for the third on the next line, and so on."

E now displays the 5th, *i.e.*, the first 60 mm. line of the series, with the remark: "Here is the first one." When *S* is ready for the second line, *i.e.*, 7–10 sec. later, *E* remarks, as he exposes it: "Here is a longer one." When the third is exposed, he remarks "Here is a shorter one;" and he continues to use these remarks, alternately, at the moment of exposure of each line, until 15 lines have been exposed, the first without suggestion, the remainder coupled with 14 suggestions—7 of shorter, 7 of longer. These suggestions must be given just before the line is exposed, in a quiet tone, without looking at *S*. *S* should see the disc turn and the new line appear at the moment that he receives the suggestion.

If desired, *S* may be questioned afterward, as indicated in Test 42, with regard to his attitude toward the suggestions.

Treatment of Data.—When *S* accepts the suggestion, record a 'plus' case; when he resists the suggestion, either by making the length equal to that of the preceding line, or by altering the length in a direction contrary to the intent of the suggestion, record a 'minus' case. The algebraic sum of these plus and minus cases may serve as an index of *S*'s suggestibility. Record should also be kept of the extent of modification (in mm.) made by *S* in each trial.

Results.—(1) The verbal directive suggestion used in this test is more potent, at least for children 8–10 years old, than the auto-suggestion induced in Test 42. Sixteen of 23 pupils tested by Binet submitted completely to the suggestion,[1] and no one resisted every suggestion.

(2) The suggestion is, in general, stronger at the outset than toward the end of the series, as is indicated by the fact that the extent of modification of line-length decreases, and the number of complete resistances increases, as the series progresses.

(3) Verbal suggestion is commonly more effective in producing augmentation than in producing reduction in line-length, in the proportion of about 5 to 4.

(4) There are marked *individual differences* in the suggestibility of school children under the conditions of this test. Binet found that in 18 trials the number of resistances to suggestion ranged from 0 to 14. (See Binet, 1, pp. 228–9, for a detailed table.)

[1] This statement is made in the text, but does not appear to be borne out by Binet's table (1, pp. 228–9).

(5) The *first line* is practically invariably underestimated.

(6) Tests upon 10 children, whose average *age* was 17 years, showed less suggestibility than in the case of younger children; still, four of the 17 accepted every suggestion, and three others resisted suggestion only once. The average extent of modification produced by suggestion is, however, less than that in the case of younger *S*'s. Again, the extent of modification is practically constant throughout in the series with the older *S*'s, but large at first and then progressively less in the series with the younger *S*'s.

REFERENCES

(1) A. Binet, La suggestibilité, Paris, 1900, especially 219–243.

(2) A. Binet and V. Henri, De la suggestibilité naturelle, in Revue philos., 38: 1894, 337–347.

TEST 44

Suggestion by illusion of warmth.—In measuring either discriminative or liminal sensitivity, difficulty is not infrequently caused by the interference of auto-suggestion (see various tests of Chapter VI). In the immediately preceding tests (Nos. 40 to 43), a process of discrimination (of weights and line-lengths) was, accordingly, made the basis for testing suggestibility. In the present test, a (supposed) measurement of liminal sensitivity is made the basis for testing suggestibility. The plan is to arrange experimental conditions in such a way as to suggest warmth, when no warmth is present.

This idea seems to have originated in the Yale laboratory, when Seashore (4), in 1895, worked out a proposal made two years earlier by Scripture (3). Small's varied tests of suggestibility (5), which appeared in the following year, embodied two very simple 'heat' tests. More recently, Guidi, in 1908, and Scott, in 1910, have reported tests of suggestibility to warmth, the former with a simple 'warmth box,' the latter with apparatus somewhat similar to the original device of Seashore. Four methods are described herewith; the resistance-wire method of Seashore and Scott, the heated box method of Guidi, and two simple methods employed by Small.

A. ILLUSORY WARMTH—RESISTANCE-WIRE METHOD

APPARATUS.—Stop-watch. Special warmth-tester.

The warmth-tester consists of a wooden box, open at the end facing *E*, and provided, on the top, with porcelain sockets for four electric lamps, wired in multiple, and with a snap switch by which the current (105–110 volt, D. C.) may be turned on or off. The wiring is purposely left visible, and leads conspicuously from the lamps to a coil of No. 24 German-silver wire, 1 m. long, which is wound, without covering, about a flat piece of hard rubber, 3 × 10 cm. This resistance coil is fastened to the front of the box, in such a manner that it may be easily reached by *S*, without exposing his fingers to the warmth of the lamps on the top of the box. A concealed circuit leads to a noiseless switch, underneath the box, which can be operated by *E* without *S*'s knowledge. By means of this switch, *E* may shunt the current through the coil, or cut the coil out entirely, without affecting the illumination of the lamps.[1]

PRELIMINARIES.—Find an arrangement of lamps such that, when the current passes through the coil, warmth becomes perceptible in 8 to 10 sec. Four 8-C. P. carbon-filament lamps generally prove satisfactory. If necessary, use one or more 16 C. P. lamps. A more powerful illumination, with the same heating effect in the coil, may be secured by the use of tantalum or of tungsten lamps.

METHOD.—Give *S* the following instructions: "I want to test your ability to perceive warmth. Hold this coil of wire gently between your thumb and two fingers, like this [illustrating]. You will see that the coil is connected with these electric lamps, so that, when I light them, a current of electricity can flow through the coil and warm it—it is made of German-silver wire, and offers a slight resistance to the current. There is nothing at all to be afraid of. You can't feel any shock from the current, nothing but a slight warmth. Watch carefully, and, the moment that you feel warmth, say 'now.'"

Without attracting *S*'s attention, close the secret coil-switch, so that no current passes through the coil. After a preliminary 'ready,' snap the lamp-switch rather ostentatiously; start the stop-watch at the same instant, and lean forward in an attitude of expectancy, keeping one hand on the lamp-switch, as if awaiting

[1] In default of a 110-volt circuit, a resistance-wire apparatus may be contrived with a battery, after the plan described by Seashore, though the absence of the illuminated lamps alters the experimental conditions.

S's 'now.' Snap the lamps off as soon as the 'now' is spoken. Record the time. Feel of the coil, or solicitously blow upon it, as if to cool it. Repeat the test 5 times with each hand, alternately.

If S, at any trial, fails to get the illusion of warmth within 60 sec., open the coil switch (without S's knowledge), so that warmth is actually felt, but record the trial as one 'resistance,' or failure.

VARIATIONS OF METHOD.—Following the plan of Seashore and of Scott, tell S that 20 trials will be made. Give a preliminary series of 5 trials with each hand, with objective warmth from the start, in each trial. Without interruption, continue with an equal number of trials in which the coil is not warmed unless S fails to report warmth within a period some 10 sec. longer than the average time at which he had reported warmth in the first 10 trials.

TREATMENT OF DATA.—In either method, suggestibility is measured by the absolute or relative number of trials (without objective warmth) in which S reports warmth.

S may also be rated in terms of the quickness (number of seconds) with which the illusion is reported.

B. ILLUSORY WARMTH—GUIDI'S METHOD

APPARATUS.—Stop-watch. Matches. Alcohol lamp, fitted with hinged extinguishing cap. Cubical wooden box, with a chimney-like metal top, a circular hole in the front face, and a hinged door in the back face (Fig. 57).

METHOD.—E's instructions are analogous to those in the resistance-wire method. "I want to test your ability to perceive warmth. I want you to thrust your forefinger into this box through the hole in front. I shall put this lamp into the box. It won't burn you at all. Just watch very carefully, and say 'now' the moment that you notice any warmth in the box." E then light, the alcohol lamp, opens the door of the box, sets in the lamps extinguishing the flame as he does so, starts the watch, closes the door, and expectantly awaits S's judgment.[1]

[1] Guidi's method deviated somewhat from the above, in that S was instructed to push his finger slowly into the box, and degree of suggestibility was measured by the extent to which the finger had been inserted when warmth was reported. This procedure presents difficulty in governing the rate of movement, and has, so far as the author's experience goes, no advantages over the procedure that has been recommended.

C. ILLUSORY WARMTH—SMALL'S METHOD

MATERIALS.—Alcohol lamp. A pin thrust through the rubber tip of a pencil, or through a small bit of soft wood. Toothpick, or other bit of wood with a blunted point. Matches. Piece of cardboard, about 15 × 15 cm. Blindfold.

METHOD.—(1) Let *S* see the lighted lamp and the pin in its holder. Instruct him as follows: "I am going to warm this pin in this flame, then touch it to the back of your hand to see if you can notice the warmth it makes. Don't be afraid of being burned,

FIG. 57. GUIDI'S APPARATUS FOR THE WARMTH ILLUSION. (Modified by Whipple.)

as it will not be hot enough for that, and I shall try it on my own hand first. Say 'now' when you feel its warmth." Blindfold *S* carefully. Go through the operation of heating the pin; say 'ready,' but do not touch *S*'s hand at all. If *S* reports warmth, ask him to describe the feeling: if he does not report warmth, repeat the test, but touch him on the back of the hand with the pointed piece of wood, to see if the contact is reported as 'warm' or 'hot.'

(2) Light a match and move it around about 1 cm. above the back of S's hand. Call his attention to the 'waves of heat' that he feels. Blindfold him carefully. Ask him to see if he can detect the heat waves every time. Strike a match, and move it about over his hand, but hold the cardboard between the match and the hand. Repeat several times with either hand. Note the number of times the suggestion is 'accepted,' and any indications of the readiness or degree of suggestibility.

RESULTS FOR ALL METHODS.—(1) In general, the results of the warmth-illusion test appear to be conditioned primarily by the success of the investigator in creating a proper atmosphere of suggestibility, rather than upon the particular apparatus employed. Thus, Seashore met with amazing success. Of his 8 college students, only 3 resisted at all, and these but once or twice each, so that, in 420 trials, there were only 5 failures to perceive heat. Small tested boys and girls from the 7th grade and the high school: in 21 trials, 5 reported heat, with no contact at all, 19 reported heat from the wooden point, while in 19 trials with the "heat-waves," 17 proved suggestible.

Of Scott's 20 college students, 9 'yielded' 10 times (of a possible 10); 5 yielded 9 times; 2 yielded 4 times, and 1 each, 8, 7, 5, and 3 times. No one of the 20 S's resisted in every trial.

Okabe, who worked with school children and adults in the Cornell laboratory under the author's direction, obtained positive results in 70.7 per cent of the trials, and with 22 of 29 S's (Table 87). The Italian children tested by Guidi were less suggestible (at least for his method), as Table 88 shows.

TABLE 87

Suggestibility to Warmth. Resistance-Coil Method (Okabe and Whipple)

GROUP	NUMBER	TRIALS	YIELDS	PERCENTAGE OF SUGGESTIBILITY	CASES WITH NO YIELDS
Men	12	59	43	73	1
Women	7	29	20	69	2
Bright boys	5	36	27	75	1
Dull boys	5	33	21	64	1
Totals	29	157	111	70.7	5

(2) The *relation to sex and to age* can not be stated with assurance. Guidi's results indicate maximal suggestibility at the age of 9, but the Cornell tests, perhaps from being too few in number, failed to show characteristic differences between grammar-school boys and adults. It is likewise unsafe to generalize from the indications there given of the greater suggestibility of men.

TABLE 88

Suggestibility to Warmth, as Related to Age. 187 Cases (Guidi)

Age	6	7	8	9	10	11	12	13	14	15
Per cent suggestible	50	40.9	51.8	62.5	50	40	33.3	21.4	27.3	33.3

(3) The *degree of suggestibility*, as indicated by the readiness with which warmth is felt, differs, as might be expected, in different *S*'s, *i.e.*, even of those who invariably perceive warmth, some report only "faint warmth," others "sudden heat," etc. Guidi classed his pupils into three groups, according as they took the suggestion quickly (in 1 to 2 sec.), moderately (2 to 3 sec.), or slowly (after 3 sec.), and found 33 per cent, 63.7 per cent, and 3.3 per cent of his *S*'s in these three classes, respectively.

(4) Scott found no *correlation* between suggestibility as measured by the warmth illusion and suggestibility as measured by his flight-of-colors test. Okabe's tests afforded the following low correlations with other forms of suggestibility tests; with progressive lines (Test 42) 0.17, with contradictory suggestion (Test 43) 0.21, with directive suggestion (Test 43) 0.29, with the weight illusions (Tests 40 and 41) none.

REFERENCES

(1) G. Guidi, Recherches expérimentales sur la suggestibilité, in Ar. P., 8: 1908, 49–54.

(2) W. D. Scott, Personal differences in suggestibility, in P. R., 17: 1910, 147–154.

(3) E.W. Scripture, Tests on school children, in Educ. Rev., 5: 1893, 52–61.

(4) C. E. Seashore, Measurements of illusions and hallucinations in normal life, in Yale S., 3: 1895, 1–67, especially 30–32.

(5) M. H. Small, The suggestibility of children, in Pd. S., 4: 1896, 176–220, especially 183–186.

CHAPTER XI

Tests of Imagination and Invention

Imagination, like most of the stock psychological terms, has the misfortune to be used in several different ways. In popular usage, imagination commonly implies something fanciful and unreal: we condemn a rumor, for example, by dubbing it "a mere figment of the imagination." In psychology, imagination has both a general and a specific meaning. Broadly speaking, imagination is equivalent to imaging, or thinking in images, as over against perceiving—re-presentation as contrasted with presentation. But the psychologist also differentiates between imaging which refers to some part of one's past experience (memory) and imaging, which, though necessarily based upon this same material, presents the material in new forms or patterns, and which is not felt to refer definitely to some part of one's past experience. This latter is imagination in the specific, or narrower meaning of the term.

A further distinction is made between imagination which occurs under passive attention, as illustrated in reverie, musing, or dreaming, and imagination which occurs under active attention, and which is marked by persistent, purposeful effort to dissociate former combinations of experience and to reorganize them into some new plan. We have, then, a distinction between passive imagination and active, creative, or productive imagination.

The five tests of this chapter are designed both to secure indications of the wealth of spontaneous imagery in phantasy, and to measure capacity for creative or inventive thinking.

In so far as intelligence denotes not merely good attention and good memory, but also inventive capacity, ability to plan and organize, to anticipate, or to "put two and two together" (Ebbinghaus' *kombinierende Tätigkeit*), in so far must the attempt to measure intelligence employ tests of productive imagination and invention. It goes without saying that the tests here described do not

exhaust the possibilities of investigation in this important field of mental activity. Undoubtedly, new tests will be devised which will prove of value in supplementing those heretofore employed. We need especially a series of tests of inventive capacity, of graded difficulty, which shall put less emphasis upon linguistic attainments.

TEST 45

Ink-blots.—In their discussion of a proposed series of tests for the examination of individual differences in mental traits, Binet and Henri, in 1895, suggested that fertility of visual imagination might be investigated by means of a series of ink-blots. Two years later, G. Dearborn published brief suggestions for making a series of blots, and in the following year described the results of the use of 120 blots in the case of 16 Harvard students and professors. Since then Kirkpatrick has tried the ink-blot test with public school children of 8 elementary grades, and Miss Sharp has followed the suggestion of Binet and Henri in a study of individual psychology upon graduate students in Cornell University.

The ink-blot test is commonly classed as a test of passive imagination, under the assumption that *S* simply looks at the blot and allows his associative processes to suggest to him whatever 'pictures' they may. In practise, however, *S* is quite likely to search actively for these associations, so that the mental activity concerned is, perhaps, more allied to active than to passive imagination.

MATERIALS.—Standard series of ink-blots, numbered from 1 to 20. Stop-watch. Paper properly prepared for recording *S*'s statements.

The primary difficulty heretofore existing in the application of the ink-blot test has been the lack of standardized material. To meet this difficulty, the author has prepared the series of blots just mentioned by using zinc-block prints, so that investigators may now apply the same series of blots, and thus secure strictly comparable data. Unfortunately, this series has not yet been applied upon a sufficiently extensive scale to render it possible to publish norms of performance for the test.

METHOD.—(*a*) *Full procedure*: Instruct *S* as follows: "I have here a series of 20 odd-shaped ink-blots. I want you to take them

in order from 1 to 20, one at a time, to look them over at your leisure, and to tell me (or write down on a numbered blank) what things you can see in each blot. Try them in different positions. Of course, these blots are not really intended to be pictures of anything, but I want to see whether your imagination will suggest pictures of things in them, just as you sometimes try to see what objects you can make out of clouds." Let *S* take his own time. Especially with younger *S*'s, it is better for *E* to record the results, so that *S* may be perfectly free to enumerate as many things as are suggested to him.

Kirkpatrick used only four blots, and allowed each pupil one minute to name as many associations as possible for each blot. Miss Sharp used 10 blots, and allowed only 5 minutes for the test.

The test may be conducted with a group of *S*'s by distributing the cards, and having them passed successively from member to member of the group until each *S* has written his associations for each card, but this method has obvious disadvantages.

(*b*) *Shorter procedure.* Following the method used by Dearborn, arrange the 20 cards face down in a pile, with the 20th card at the bottom, the 1st at the top, and the numbered edges toward *S*. Instruct *S* as follows: "Each of these 20 cards has on it an odd-shaped ink-blot. When I say 'now,' turn over the first card in this way [illustrating the movement that will expose the face of card No. 1 with the numbered edge toward *S*]. Look at the ink-blot, without turning the card in any other position, and say 'now' (or tap on the table) as soon as you have thought of something that the blot resembles. Of course, the blot is not really intended to be a picture of anything, but I want to see whether your imagination will suggest some 'picture' in it, just as you sometimes try to see what object you can make out of a cloud." Give the command 'now'; start the stop-watch at the same time. When S gives his signal, stop the watch, record the time and the object or association given by *S*. Continue in the same manner with the remaining cards.

Treatment of Data.—In the full procedure, the score is based upon the average or total number of associations: in the shorter procedure, upon the average speed of the single associations. It is also possible to form some estimate, in either case, of the type,

richness and variety of *S*'s imagery by classifying the associations after some such plan as that illustrated below from Miss Sharp's results.

Typical Results.—The following associations for the 20 cards of the standard series are taken from the records of three adults, and will serve to indicate the variety that may be expected when the records of several *S*'s are compared.

(1) A lady seated on a couch. A witch riding on a new moon across the sky. A moose's head. A woman, sitting on a bank of shrubs, waving a handkerchief.

(2) Child, crouching in fear. Man with grotesque features. Ugly old colored woman.

(3.) A banner. A right-angled triangle. The God Billiken. An Egyptian idol. A jade-stone idol.

(4) A large beetle. A boat load of excursionists. A lobster. A spider. Potatoes. A dirigible balloon of the Zeppelin type, with a cloud of steam or smoke overhead, and a grappling-anchor trailing below.

(5) A pig. A woman with a big head of hair. A butterfly. A hole through the ice. A girl wearing a tam-o-shanter cap. Human liver and heart. A rock.

(6) Woman running and holding her skirt. Woman with a muff in her left hand, and her hat almost blown off. A broken bellows. Merry Widow waltz.

(7) Large caterpillar on a horse's shoulders. A devil bending over something. An old man. A dream monster.

(8) Human torso. Hot and cold water faucet in a bath tub. Person with head bent forward, holding sticks in her hand. Heads of two birds trying to swallow what is between them.

(9) A goat with a pack on his back. A turkey with drooping wings trailing on the ground.

(10) Ugly man's head. Head and arm of a woman with a lighted candle in her hand. A dachshund running off with some one's cape.

(11) Map of Scotland and Ireland. Owl that has just placed a fish before him on the branch of a tree. Some specimen in geology. A tree blown in a heavy gale.

(12) Map of United States and part of Canada. A chicken lying on its back. An Indian head. A woman sitting on a cliff under a tree, reading a novel.

(13) A flying squirrel. The skin of a bear. A hen sitting on a nest.

(14) A crab. A bat with outspread wings. A moth. A neurological slide.

(15) Section of medulla oblongata. Two nuns bowing their heads together. A tulip. A false mask.

(16) Closed hand with thumb and little finger, or a sixth finger, projecting. A loving cup. A tea-pot.

(17) A root. A porcupine.

(18) A Chinese dragon, as seen on packages of fire-crackers. Branch of a gnarled oak.

(19) Bird alighting on a nest. A flying squirrel.

(20) Man pulling off his sweater. Runner leaning forward to start a foot-race. Photographer, with focussing cloth over his head. Crocodile suspended by the head. Bear with the grandmother's night-cap and gown, as illustrated in Little Red Riding Hood.

General Results.—(1) *Speed of association.* In 1920 trials, Dearborn found the average time for making a single association to a blot to be 10.3 sec. This seemingly long time may be due to the difficult nature of some of the blots in his series.

(2) *Dependence on age.* Kirkpatrick states that "younger children seemed more suggestible or imaginative, as they named more spots." This conclusion is based upon the data presented in Table 89.

TABLE 89

Average Number of 'Names' given to Ink-Blots (Kirkpatrick)

GRADE	I	II	III	IV	V	VI	VII	VIII
Average	2.9	2.5	2.6	1.8	1.9	1.7	2.1	2.2

It is evident that something besides a simple decline of 'imagination' with age is exhibited in this table. In explanation, Kirkpatrick says: "The younger children seemed to have no doubt whatever of the spot being a picture of the object they named, while the older children simply said 'it is some like' or 'it looks a little like, 'a dog,' 'cloud,' or whatever else was suggested. This superiority of the small children is striking when we consider that the number of mental images that they have is much smaller than that possessed by older children, who may name a part of the body or the map of a country or something else that the younger children know nothing about.

"The smaller number of objects seen in the spots by the children of the 4th, 5th, and 6th grades is probably to be explained by the fact that children of those ages have become more critical in their sense-perception, as their ideas have become more definite, and as they have learned from life's experiences and from training to be more careful in their judgments. The older pupils of the 7th and 8th grades, on the other hand, have passed into another stage in which they realize that a picture is not necessarily this or that, but may resemble any one of several things, hence they are not afraid to say what it looks like."

(3) *Dependence on occupation.* Dearborn believes that, at least in maturer *S*'s, the results of the ink-blot test are conditioned, not so much by age or sex directly, as by habits of living, occupation, and other environmental factors: thus, we should expect characteristically different results from the test when applied, for example, to artists, farmers, laborers, professional men, to the city-bred or the country-bred, etc.

(4) *Individual differences,* both in speed, number and type of association seem to have been clearly marked and fairly constant, whenever the test has been applied. Thus, in Dearborn's single-association method, the highest agreement in the answers of his *S*'s for any one card was but 40 per cent, while for several cards, no two *S*'s gave the same answer.

As regards fertility of imagination, Miss Sharp noted that the most imaginative *S* in her group saw 81 objects, the least imaginative but 27 objects in the same 10 blots. The same investigator believes, however, that all *S*'s might be roughly divided into two groups, (*a*) the constructive or imaginative, who put together concrete details "in such a way as to form a significant whole," and (*b*) the matter-of-fact, or scientific type, given more to analysis than to creative synthesis.[1]

As examples of this difference, the following reports from two of Miss Sharp's *S*'s may be quoted: both refer to the one blot.

(1) *Associations few and non-constructive.* "An eagle. Stuffed turkey. Head and neck of a musk-rat."

(2) *Associations numerous and constructive.* "Giraffe. Prehistoric bird in flight. Fairy riding on a bumble-bee. Bit of tropical jungle, with trailing gray mosses and pools of water. Japanese lady. Bit of landscape with two hills and a valley between—an army encamped under one hill. Moss-grown log floating in water. Fabulous monster (griffin perhaps), walking off on his hind legs with a small Hottentot under his arm."

(5) *Qualitative classification.* It is often possible to classify the associations peculiar to a given *S*. Thus, Miss Sharp mentions as classificatory groups: (*a*) common-place, every-day objects, such as domestic utensils, tools, plants, and particularly animals, (*b*) scientific objects, such as geometric figures, schematic drawings,

[1] It is tempting to regard this classification as identical with the common classification of laboratory *S*'s into 'subjective' and 'objective' observers.

(*c*) objects suggested by literary reminiscence, and (*d*) objects from fable and mythology, such as centaurs, dragons, witches, fairies, etc. Some *S*'s exhibit variety of association, in that they cite objects that belong to several of these groups; others are much less fertile in imagination and confine themselves largely to a single type of imagery.

REFERENCES

(1) A. Binet and V. Henri, La psychologie individuelle, in A. P., 2: 1895 (1896), 411–465, especially 444.

(2) G. Dearborn, Blots of ink in experimental psychology, in P. R., 4: 1897, 390–1.

(3) G. Dearborn, A study of imaginations, in A. J. P., 9: 1898, 183–190.

(4) E. Kirkpatrick, Individual tests of school children, in P. R., 7: 1900, 274–280.

(5) Stella E. Sharp, Individual psychology: a study in psychological method, in A. J. P., 10: April, 1899, 329–391.

TEST 46

Linguistic invention.—The ink-blot test serves primarily as a test of visual imagery. But an even more fruitful source of individual differences in creative ability may be found in linguistic invention. Miss Sharp, acting upon the suggestions of Binet and Henri, tested what she terms 'literary imagination,' in three ways, viz: by the development of sentences, by the development of a given theme, and by the choice of a topic for composition. The last-named test seems, for the present purposes, to have relatively slight value: on the other hand, the construction of sentences may be tested either, as was done by Miss Sharp, by assigning three words which must be incorporated into a sentence, or, as was done later by Binet, by giving *S* partially written sentences to be completed.

Before undertaking these three formal tests, however, it is desirable, if the purpose in mind is to make a qualitative study of the mental type of individual *S*'s, to institute a preliminary inquiry concerning the general literary tastes and habits of each *S*. The exact nature of this inquiry must, naturally, be adapted to the age and

training of the S's: the following are some of the points that have been covered by investigators: (1) list of favorite books, (2) statement of favorite type of reading, (3) statement of the magazines, periodicals, newspapers, etc., ordinarily read, (4) list of books (outside of classroom or professional work) read during the last year, (5) statement of favorite games and evidence of enjoyment of games, like chess and checkers, that demand creative activity and foresight, (6) fondness for the theater, drama, music, painting and other forms of art, etc.

A. DEVELOPMENT OF SENTENCES

METHOD.—Ask S to write as many sentences as possible containing the three nouns: *citizen, horse, decree.* Each sentence must contain all three nouns, though it may contain others as well. The sentences are to be as varied as possible. Five minutes are allowed. Continue the test with four more sets of nouns, and afterward make similar tests with five sets of verbs. For the noun tests, use as additional sets: (2) *bell, ground, owner,* (3) *skill, modification, picture,* (4) *cup, fraction, money,* (5) *letter, law, summer.* For verbs use (1) *bless, destroy, write,* (2) *make, correspond, remain,* (3) *require, choose, run,* (4) *see, find, throw,* (5) *remember, put, depart.* In the noun tests, S is permitted to use either singular or plural forms, and possessive as well as nominative or objective cases: in the verb tests, he may use any form of the given verb, *e.g.*, *blessed, to bless, will bless,* etc., as well as *bless.*

The tests may be conducted with individuals or with groups, but it may be preferable, especially with young S's, to work individually and to let S dictate the sentences instead of writing them.

TREATMENT OF DATA.—The quantitative score is determined by averaging the number of sentences written by S. The quality of work may be graded upon any convenient scale, *e.g.*, 1 to 5, corresponding to five degrees of excellence. Miss Sharp used the symbols A, B, and C, and indicated intermediate grades by the use of $-$ and $+$. For purposes of computation, she then assigned numerical values to these symbols, as follows: $A- = 40$, $A = 50$, $A+ = 60$, $B- = 80$, $B = 100$, $B+ = 120$, $C- = 160$, $C = =200$, $C+ = 240$. In practise, this scoring is virtually equiva-

lent to estimating quality of work in terms of average number of words per sentence, and this simpler method may be used for the qualitative score.

Typical Results.—(1) The following are selected single sentences reported by Miss Sharp for the first test:

1. "Decrees are made for citizens, not for horses." (The connection of the words here is simple and mechanical.)

2. "That stalwart citizen on the great gray horse is a man to be trusted with the decree." (This implies a concrete situation.)

3. "All the well-to-do citizens of the village, each mounted on a horse, rode through the streets, proclaiming their dissatisfaction with the new decree." (A situation is here more fully outlined.)

(2) The following is a full set of sentences written by a graduate student, in 5 min., for the first assignment:

1. A decree was posted that the citizen should not abuse the horse.
2. The horse of the citizen was sold by official decree.
3. "Here," said the citizen, "is the horse mentioned in the decree."
4. Early in Arabian history, a decree raised to a higher caste, a citizen who owned a horse, but later, possession was sufficient for better standing, and the law was not needed.
5. If a citizen keep a horse, it is a decree that he use it kindly.
6. "What a funny decree," exclaimed the citizen, when he read of the horse sun-bonnet law.
7. The decree was signed that the horse had kicked the citizen, and therefore the injured man could collect damages from the owner of the animal.
8. "Time is up," cried the citizen, stop-watch in hand, "I hereby decree that you write the word horse and stop at once." [Faulty on account of the use of 'decree' as a verb.]

(3) The following represent groups of sentences written for the author by two college students (selected at random from a number of papers) for the fourth set of verbs. The relatively greater variety of the second group is clear.

A. 1. "I saw the book and tried to find a place in which to throw it."

2. "I threw the cat in the creek and turned to see if anyone had found me out."

3. "I see that I can find nothing to throw at him."

4. "You see, it was this way, I simply found the hatchet and threw it."

B. 1. "The child saw a horse, found a stone and threw it at him."

2. "When you find a clover, see if it has four leaves: if not, throw it away."

3. "Throw the paper out of the window and see if it will find a good landing place."
4. "Find me a pencil, then I will see if I can find out the solution to the problem which is on the paper that you threw into the basket."
5. "The boy found an apple, but when he saw it was decayed, he threw it away."

CONCLUSIONS.[1]—(1) *Dependence on part of speech assigned.* All *S*'s tend to write fewer, but better sentences with verbs than with nouns.

TABLE 90

Scores of Seven Adults in Developing Sentences (Sharp)

FORM OF TEST	QUANTITY OF WORK			QUALITY OF WORK		
	Average	Maximum	Minimum	Average	Maximum	Minimum
Nouns 'given'	4.6	6.6	3.2	79	113	55
Verbs 'given'	3.8	5.8	2.5	93	133	54

(2) The rank of *S*'s, both in quality and quantity of work, is the same when nouns and when verbs are assigned.

(3) "In general, the subjects who made the most sentences made the most elaborate, and those who made the fewest sentences made also the simplest and most unimaginative."

(4) This test correlates with the ink-blot test, in so far as those *S*'s who show most constructive capacity with the blots also show most constructive capacity in the development of sentences.

B. COMPLETION OF SENTENCES

MATERIALS.—Printed forms containing 25 beginnings of sentences,[2] with spaces for the completion of each sentence. Piece of white cardboard. Stop-watch.

METHOD.—Give *S* the following instructions: "On this paper there are printed the beginnings of a number of sentences. I am going to show these to you, one at a time. As soon as I show you

[1] These are all drawn from the work of Miss Sharp.

[2] The first 20 sentences are taken, with such slight modifications as translation has suggested, from Binet. The last five (since Binet prints but 20 of the 25 he recommends) have been supplied by the author.

one, I want you to write the rest of the sentence. You can write anything you want, as long as the whole sentence will make sense when you have finished it." If *S* fails to understand what is wanted, supply him with an extra paper in which a few trial sentences have been written in pen and ink, and show him how they might be completed. For the test proper, cover the entire test-blank with the cardboard: after a warning 'ready,' expose the first incomplete sentence.[1] Start the watch at the same time. Record as nearly as possible the time used by *S* in starting to complete the sentence, *i.e.*, the time he takes, after he reads the sentence, to 'get an idea.' The timing should be done without *S*'s knowledge.

TREATMENT OF DATA.—Compute the average, or determine the distribution of the times needed by *S* to start the 25 sentences. For a qualitative index, estimate as well as possible (preferably by using some such system of scoring as that described in the development-of-sentences test) the general value of the completed sentences. *S*'s sentences may also, if desired, be classified in regard to type, *e.g.*, vague or meaningless, commonplace, reminiscential, imaginative, aphoristic, etc.

RESULTS.—(1) Binet found characteristic differences in the *speed of work* of his two daughters, Armande and Marguerite. Thus Armande's records show 12 sentences started in less than 5 sec., 4 sentences in from 5 to 10 sec., 6 in from 10 to 20 sec., one in 28 sec., and one in 70 sec. Marguerite's records, on the other hand, show but one sentence started in less than 5 sec., but 7 sentences in less than 10 sec., and the remainder in much longer times, *e.g.*, 20, 50, and 70 sec.

(2) Binet's two *S*'s also showed characteristic differences in the *type of sentence-completion:* Armande is poetic and imaginative; Marguerite's sentences are more precise, more practical, more in accord with real life, less emotional. For example, for Sentence 1, Armande writes: "I entered the field by a covered footpath." Marguerite writes: "I entered a grocery and bought two cents worth of chocolate."

[1]The sentences have purposely been numbered from the bottom of the page so that the cardboard will not interfere with *S*'s writing.

C. DEVELOPMENT OF A THEME

METHOD.—Supply *S* with writing materials, and give him 10 min. (or perhaps longer if working with young *S*'s) to write upon some theme selected from the following: (1) *The Death of a Dog*, (2) *The Capture of a Fortress*, (3) *The Escape of a Prisoner*, (4) *A Forest Fire*, (5) *The Mission of Music*, (6) *The Influence of Newspapers*, (7) *The Delays of Justice*, (8) *A Trip in a Flying Machine.*[1]

TREATMENT OF DATA.—Quantity or speed of work may be reckoned with approximate accuracy by counting the number of words written in the assigned time: quality of work, which is equally important, especially in the treatment of imaginative themes, must be estimated by *E* after a trial has shown what may be deemed poor, and what good work for *S*'s of the age under investigation. Quality may be recorded in the manner already described, or upon the basis of 100, as in grading school compositions.

RESULTS.—(1) The relative number of ideas elaborated by different *S*'s is indicated with fair approximation by the relative number of words written, so that number of words may stand as a fair index of fluency of ideation and general linguistic readiness.

(2) As a rule, more words are written upon imaginative, than upon expository themes. Sharp's best *S* wrote in 10 min., on an average, 259 words upon imaginative, and 222 upon expository themes; her poorest *S* wrote, on the average, 124 and 94 words, respectively, for the same types of themes.

(3) Those *S*'s that show constructive ability in the ink-blot test, and in the development of sentences, also exhibit the same superiority here in the development of themes.

NOTES.—These tests of linguistic invention might, without great difficulty, be paralleled in other fields of constructive effort. A test of musical ability (of the creative sort) might, for example,

[1] The first of these themes was used by Binet in his comparative study of the mental processes of his two daughters: the next six were used for a similar purpose in Miss Sharp's study of university students—the first three of them being designed to involve imaginative, the second three expository treatment: the last theme is suggested by the author as more suitable for younger *S*'s. To secure a more reliable estimate of *S*'s efficiency it is desirable that more than one theme should be developed.

This test lends itself readily to group treatment, since it involves a familiar type of school activity.

be devised by asking *S*'s to finish a partially given musical theme, or to construct a simple melody from a given series of notes. Similarly, certain forms of artistic invention might be tested by asking *S*'s to sketch designs for wall-paper or patterns for Venetian iron-work.

Miss Sharp's test of the choice of a theme was conducted by asking *S*'s to select, from the following 10 themes, those five upon which they would prefer to write, if asked to do so: A. Imaginative themes, (1) *In a Snowstorm*, (2) *A Polar Landscape*, (3) *A Puritan Sabbath*, (4) *My Opposite Neighbor*, (5) *Man Endowed with the Power of Flight:* B. Expository themes, (6) *Civilization not Regeneration*, (7) *Wisdom in Charity*, (8) *Friendship of Books*, (9) *Fiction as a Vehicle of Truth*, (10) *The Eloquence of the Bar and that of the Pulpit.* The expository themes were generally preferred, but some *S*'s, who, as others tests showed, had little capacity to handle imaginative themes, didselect several from this division.

REFERENCES

(1) A. Binet and V. Henri, La psychologie individuelle, in A. P., 2: 1895 (1896), 411-465, especially 444.

(2) A. Binet, L'étude expérimentale de l'intelligence, Paris, 1902. Pp. 309. Especially ch. x.

(3) Stella E. Sharp, Individual psychology: a study in psychological method, in A. J. P., 10: 1899, 329–391.

TEST 47

Word-building.—The word-building test was suggested by the familiar game of anagrams as well as by the advertisements often seen in magazines in which a prize is offered to the person who can make the most words from a given word or series of letters. This test is easily administered and evaluated: it is one that calls for ingenuity and active attention: it might fairly be said to demand that ability to combine isolated fragments into a whole, which Ebbinghaus has declared to be the essence of intelligence and for the measurement of which he devised his well known 'completion method;'[1] and finally, its execution is conditioned to a certain extent by the richness and readiness of the examinee's word-vocabu-

[1] See Test No. 48.

lary. One may expect, therefore, to find a correlation between this test and the vocabulary test (No. 50) and possibly between it and school standing or general intelligence, and other tests of creative literary ability.

Materials.—Two specially prepared blanks, the first of which calls for combination of words from the letters *a, e, o, b, m, t,* the second from the letters *e, a, i, r, l, p.*

Method.—Provide *S* with the first test blank, and give him the following instructions: "Make as many words as you can from the six letters given on this blank. You may use any number of letters from one to six, but no letter may be used twice in the same word, and no other letters than these six are to be used. You will have five minutes." Conclude the test by use of the second blank under the same conditions.

Results.—The author has elsewhere (2(reported the results of the use of the first test-slip with two groups of college students and a group of 50 boys from the 7th and 8th grades of the Ithaca public schools, and of the second test-slip with the same college students. From this report the following statements are drawn:

(1) As will be seen in Table 91, the second slip offers more possible combinations, so that more words and a greater number of different words are made with it.

TABLE 91

General Results in Word Building (Whipple)

Test	Grade	Male	Female	Largest	Smallest	Average	Mean Variation	Total No. Different Words	Possible Words
1	College	7	15	25	10	18.6	3.54	41	70
1	"	9	27	26	10	18.6	3.30	41	70
1	Grammar	50	0	21	6	12.4	2.37	38	70
2	College	7	26	32	14	21.9	4.00	57	105
2	"	9	27	33	15	23.5	3.66	61	105

(2) The individual differences in rank are large: thus it happens that not a few grammar-school boys make more words than some of the college students: to be more explicit, it may be stated that, with the first test, 10 grammar-school pupils make 15 words or over, while 13 college students make fewer than 15 words.

(3) By examining the papers in detail, and tabulating the total number of words formed and the number of times each of these words is given, one may discern something of the principles which govern the operation of the test. The following are the data thus secured:

Test No. 1. 58 College Students. (45 Different Words.)

Over 50 times—bat, mat, bet.
40–49 times—eat, met, Tom, at, boat.
30–39 times—meat, to, tea, beat, team, tab, ate, am, moat, mob, me, beam, toe.
20–29 times—tame, boat, be, mate.
10–19 times—boa, mote, bate, abet, tomb, tome, tam.
5– 9 times—Mab, Abe, Mae, ma, atom, a.
1– 4 times—bot, mot, o, Moab, beta, bema, tabe.
Not given[1]—ab, ambe, ambo, amt, atmo, ba, bam, bo, bom, boma, bote, ea, eam, eb, em, eta, mao, meta, mo, moa, moe, ob, obe, om, ta, tambo, tema.

Test No. 1. 50 Grammar-Grade Boys. (38 Different Words.)

Over 40 times—mat, bat.
30–39 times—bet, at, met.
20–29 times—to, eat, Tom, beat, tea, meat, be, am, boat.
10–19 times—toe, mob, beam, me, ate, team, tab, boa, oat.
5– 9 times—ma, bate, a, moat, mot, tame, mate, bot.
1– 4 times—tam, tomb, Abe, mote, Moab, Mae, o.
Not given—those not given by college students, plus abet, atom, bema, beta, Mab, tabe, tome.

Test No. 2. 69 College Students. (66 Different Words.)

Over 60 times—lip, lap.
50–59 times—rip, rap, pear, ear, real, pie, leap, rail, pale, reap.
40–49 times—reap, pail, pile, ale, pair, are, ape, lie, pea, peal.
30–39 times—pare, earl, pearl, air, par, lair, ripe, liar.
20–29 times—ail, Lear, rape, ire, pal.
10–19 times—lea, pa, rile, pire, era, pier.
5– 9 times—per, a, alp, Eli, plea.
1– 4 times—I, paler, peril, lira, rep, rale, ile, lare, ra, pil, piler, ril, April, Ira, la, pareil, pi, pilar, Rea, Rae.
Not given[1]—ai, aiel, aile, aire, al, apl, ar, Ariel, aril, ea, el, ela, epi, er, eria, il, irp, le, lep, lepra, lerp, li, lier, lire, lirp, paie, pali, parel, parl, pela, pel, pia, piel, pila, plie, plier, prial, prie, re, rei, rial, ril, ripa.

[1]This list is based on the words actually given in the Standard Dictionary, not including, however, Scotch terms.

Inspection of these lists shows (*a*) that three-letter words are in every instance those most frequently formed, (*b*) that two-letter words and the one-letter words, which one might expect to be most frequent since most simple, stand relatively low, *e.g., ma, be, am, pa, me, a, o, I*,[1] (*c*) that grammar-school boys give all the words given by college students save a few rather unusual terms such as *atom* and *tome*, (*d*) that usage and ordinary speaking vocabulary condition the formation of words, in as much as the most ordinary words have the greatest frequency, *e.g., bat, mat, bet, eat, lip, lap*, whereas words that are less frequently used in every-day speech, although their meaning is doubtless perfectly well known, do not suggest themselves so readily under the conditions of the test, *e.g., tomb, tome, era, plea, paler*, (*e*) that the words not given by any one are, with one or two exceptions, *e.g., plier*, words of extremely rare usage or unusual form, alternative spellings, etc.

Terman has pointed out still other factors that condition the outcome of the test. "Much depends, of course, upon the vocabulary at command, and this in turn depends largely upon home training and amount of habitual reading as well as upon native retentiveness. A second factor is ability to spell, and habits of word analysis generally. Very important, also, is the use of a rational plan; some skipped about and made combinations at random, while others took the letters one by one and joined them in as many different ways as possible with the others. Lastly, the rate of shifting of attention, and the degree of mental inertia as opposed to spontaneity, also contribute to the result" (1, p. 342).

(4) *Sex differences* may not with certainty be made out, yet, as Table 92 indicates, in every group tested, the men did slightly better than the women: though this difference is small—less than the mean variation, yet it is constant in direction in all four instances.

(5) In the case of grammar-school pupils no *correlation* could be established between word-building and class standing: in the

[1] It appeared, upon inquiry, that some of the college students had omitted words like *pa, ma, a, o*, and *I* on the ground that they were 'not real words,' or 'didn't count,' but, oftener, they seem to have been passed over because the attention was concentrated upon the making of *combinations*.

case of the 58 college students there appeared the insignificant correlation, $r = +0.13$, P. E., 0.08. Terman, however, found his stupid boys generally inferior to his bright boys.

TABLE 92

Sex Differences in Word-Building (Whipple)

TEST	DATE	MEN	AVERAGE	WOMEN	AVERAGE
1	1906	7	18.7	15	18.6
1	1907	9	19.7	27	18.0
2	1906	7	23.6	26	21.4
2	1907	9	25.8	27	22.7

REFERENCES

(1) L. M. Terman, Genius and stupidity: a study of some of the intellectual processes of seven 'bright' and seven 'stupid' boys, in Pd. S., 13: 1906, 307–373.

(2) G. M. Whipple, Vocabulary and word-building tests, in P. R., 15: 1908, 94–105.

TEST 48

Ebbinghaus' completion method.—In July, 1905, the school authorities of Breslau requested certain persons, among them Professor H. Ebbinghaus, to undertake a scientific investigation of the fatigue-effects of the continuous five-hour session then in vogue in that city. In the course of this investigation, Ebbinghaus devised and applied, in conjunction with other tests, what he termed the '*Combinationsmethode*' (since referred to by Elsenhans (2) as the 'completion-method' and by others as the mutilated text test).[1]

[1] Meyer (6) has pointed out the inaccuracy of the translation "combination-method," which has been current for some time. The German *Combinationsgabe* is not a talent for combination, but an ability to "put two and two together," or, to use Meyer's explanation, "a talent for drawing conclusions from premises which do not very readily present themselves to a man's consciousness as items of a unitary logical thought, but which, *as soon as they are combined*, suggest the conclusion very forcibly." This is quite true, but the author can not see that Meyer has improved matters by advocating the translation "conjectural method." To conjecture is to surmise, to guess, to form a tentative opinion, inferentially. Technically, the activity in the Ebbinghaus test might be labelled 'redintegration,' but, as this term is somewhat clumsy, the designation 'completion method' seems entirely adequate.

The author of the method says in substance: Mental ability demands not merely retentive capacity, readiness of recall, or facile association of specific past experiences; it demands all this and something more, something more complex and, as it were, creative; namely, the ability to combine, into a coherent and significant whole, mutually independent and even seemingly contradictory impressions. In short, intelligence is essentially a combinative activity. To measure intelligence, therefore, we must employ a test that demands ability to combine fragments or isolated sections into a meaningful whole. Such a test may be afforded by mutilated prose, *i.e.*, by eliding letters, syllables, words, or even phrases, from a prose passage and requiring the examinee to restore the passage, if not to its exact original form, at least to a satisfactory equivalent of it.

On account of the enthusiastic statements of Ebbinghaus, who characterizes this method as "a real test of intelligence," and as "a simple, easily applied device for testing those intellectual activities that are fundamentally important and significant both in the school and in life," the test has assumed some prominence. Thus, Wiersma used it to determine the relative force of sex, age, native talent, and school training as determinants of the mental efficiency of his pupils; Terman used it in his study of stupidity and genius; Krueger and Spearman embodied it in their investigation of the correlations between different phases of mental capacity, while Lipmann and Wertheimer expanded it into a diagnosis-of-fact test (*Tatbestandsdiagnostik*). Its usefulness as a measure of fatigue, which has been questioned especially by Kraepelin, has, however, been more or less lost sight of since its initial employment for that purpose by Ebbinghaus.

The completion method is peculiarly difficult to class psychologically, for the simple reason that the nature of the mental processes that it demands depends almost entirely upon the number and kind of elisions that are made in the text. To take extreme cases, if the elisions are numerous and sweeping, it may become really a linguistic puzzle of a very difficult variety, and it then belongs rather in the group of tests of active or creative imagination of the literary type: if, on the other hand, the elisions are but few and simple, it may degenerate into a simple test of controlled

association of any desired degree of ease. Again, if the mutilated text be first read to the examinee, as some, *e.g.*, Elsenhans, suggest, the test becomes in the main a test of associative recall, *i.e.*, a form of memory test.

Again, since the elision of a single letter may, in some circumstances, very considerably increase the difficulty of the test, it follows that, without extensive preliminary trials, it is well-nigh impossible to prepare a series of texts of equivalent difficulty, or to insure that the several sections within a given text present equivalent difficulty.

The following is a sample section of text as used by Ebbinghaus and other German investigators: the dotted lines indicate the position and approximate length of the omissions.

Belagerung Kolbergs. 1807.

"Da der Feind fortf...... an .... neuen Schanze am Sandwege.... angestr....... Eifer zu .... so hatte unser neuer Kommandant gleich ... . ersten Nacht .... Hierseins einen Aus......... .... dieselbe angeordnet," etc.

Terman elided, in the main, whole words, instead of syllables, on the ground that the word is a more natural unit of language than the syllable, and that ability to supply missing syllables will, in the case of school children, depend largely on the extent to which word-analysis has been taught in the schools: this varies in different school systems and even in different classes of the same system. He employed one text with, and one without preliminary reading. Since these are the only English texts that have been employed, they are prescribed for the test. There has also been added another text (No. 1), arranged by the author in accordance with Ebbinghaus' plan of eliding portions of words as well as entire words.

MATERIALS.—Stop-watch. Three printed texts. [If all three texts are to be used for the test, *E* should prepare a short sample piece of mutilated text, say three or four lines, which may be typewritten, or placed on the blackboard for group work, and used for demonstration and preliminary trial. If one of the texts is not used, this may serve the purpose.]

In the reproductions here given, italics indicate the elisions in the printed texts. Text No. 1 contains 100, Text No. 2, 93, and Text No. 3, 100 elisions.

Text No. 1.

Where the Dandelions Went.

Wh*en* Willy *was* two *years* old, he *lived in a* red farm-h*ouse with a* yard *in* front *of it.* The dan*delions* were *very* th*ick* there; so that *the yard* lo*oked* yellow instead of *green.*

One bright *day* Willy's m*amma* put *on his* straw *hat and* sent him *out into the* yard to *play.* She knew *the yard* had *a* high *fence;* and *he* could not o*pen the* gate; so he *was safe.* W*hen* it *was* time *for* him *to have a* nap and *she* went *to call* him, she noticed that *a* gre*at* m*any* of the da*ndelions were gone.* She won*dered* where *they were;* but, as *Willy cou*ld not talk much, *she* did not *ask* him *about* them.

A short *time* after, while *he* was *asleep in* his crib, his mamma went *out to draw some* wa*ter.* When *the* buck*et* came *up* full *of water, the* top was all *yellow* with *dandelions.* Look*ing* down into *the well, she* could *see* no *water* at all, on*ly dandelions.*

It was no wonder, then, where *the* bloss*oms* had *gone.* Willy *had* been *very* busy, try*ing to* fill *up the well.*

Text No. 2.

The Strength of the Eagle.

One *day the* eagle *went* with the *other* birds *to* see *which* could *fly the* highest. *They* agreed *that* he who *could* fly *the highest* should *be* called *the* strongest *bird.* All started *at the* same *time* and *flew* away among *the* clouds. One by *one* they *grew* weary *and* re*turned,* but *the* eagle *flew* upward and *upward* until *he* was *a* mere speck *in the* heav*ens.* When he *came* back, *the* others were *waiting* for him; and *when he* touched *the ground* a linnet *flew* off *his* back where *he had been* hidden and *said* that *he* himself *was the* strongest *bird.* "*I am* stronger *than the eagle,*" said the *linnet,* "for not *only* did I *fly* as high, but *when* he began *his* downward *flight,* I *left* my hiding *place* and *flew* up *a* little *higher.*" *At* this boastful *speech* the *others shook* their heads and *called a* council to *decide* the matter. After *a* long *debate they* decided *that* the *eagle was* the *strongest* bird, *for* not only *did* he *fly* so high, but *he carried* the *linnet* as well.

To *this* day *the* plumes *of the eagle* are emblems of str*ength* and cour*age.*

Text No. 3.

Why the Mole is Blind.

An *Indian* once *chased a squirrel* into cloud*land.* Then *he set a* trap *for him,* laughing to *think* how *he* would *catch him.* The *squirrel* did *not come* back, but *alas!* the *sun* on *his* daily *rounds fell* right *into the trap.*

When *the* bright *sunlight did not* come, the *Indian* began *to* be *uneasy,* and when *he* found *his trap* had *the sun* fast he *did not* know *what to* do.

He tried *to get* near enough *to loosen* the cords, but *the heat* from the *sun scorched* him and he gave *it up.*

Then *he coaxed* many *animals* to try *it*, but *they* all found *the sun* too *hot*. At *last* the *mole said:* "I will *dig* through *the ground* under the *trap* and so get at *the cords*."

This *he did* and the *sun leaped* up *into the* heavens.

But it *went* so *quickly* that *the* poor *mole could* not *get away*, and the *heat* of *the sun put* out *his eyes*.

Since then the *moles* have had *to live in* dark *places* and unless one *looks* very *closely* he *can not* find *their eyes*.

Method.—Provide *S* with a demonstration or practise text (either one of the three regular texts not to be used subsequently, or the special sample prepared by *E*). Explain the nature of the test, in accordance with the directions printed on the test-blanks. It is well, in addition, to suggest that, in case a certain elision offers special difficulty, it may be temporarily passed by, since the correct interpretation of the context further on will often give the needed cue for the omitted elision.

When it is clear that *S* understands the conditions, proceed with the test proper. Use either Text No. 1 or Text No. 2, or both successively, and record *S*'s time with the watch.

Make notes of the manner in which *S* undertakes the test. Does he read it all over at first? Does he work systematically? Attentively? With confidence or hesitation? Does he grasp the general thread of the story?

Variations of Method.—(1) To conduct the test with the memory feature, employ Text No. 3, which is specially devised for that purpose. After the preliminary trial, read the unmutilated text for No. 3, entire, to *S*. Then supply him with the No. 3 test-blank and proceed as before. The text may be read more than once, or any desired time-interval may be introduced between the reading and the execution of the completion. Other variations will suggest themselves, *e.g.*, auditory, visual, or auditory-visual reading, etc.

(2) For group tests, use either of the methods just described. Limit the time so that the fastest *S* in the group tested can but just complete the work. For adults, seven minutes may be employed for the first two texts, a shorter time for the third. Group tests, for the reasons already given, are not to be recommended: it is better to allow each *S* to complete the entire test.

Treatment of Data.—In the individual form of procedure, rapidity of work may be measured in terms of elapsed time, accuracy by the percentage of 'completions' that 'make sense' and that do not violate the instructions. For net efficiency, speed and accuracy might be reduced to a single numerical symbol, after the methods suggested in the Cancellation Test. For mature *S*'s, working individually with the texts here used, the quality of work will commonly be very high, so that speed of work itself may be used as an index of efficiency.

As used by Ebbinghaus, Krueger and Spearman, and others, however, the time-limit method has been followed, with an allowance, for the German texts, of 4 or 5 minutes. In this case, quantity and quality of work are computed as follows: (1) Give a credit of 1.0 for each elision filled in in any manner. (2) Give a debit of 0.5 for each elision unfilled in any manner. (3) Give a debit of 1.0 for each elision filled in such a manner as not to make sense, or for each word introduced in excess of the number called for by the lines that indicate elisions (or for each word that is quite obviously too short or too long for the space assigned for completion, even though the passage 'makes sense'). For quantity of work done, add (2) and (3) and subtract the sum from (1). For quality of work done, compute the relation in per cent of the same sum to (1).

Typical Results.—The following is a sample of the work of a boy, 11 years old, one of Terman's "bright" group, who 'completed' the second text, with the exception of three elisions, in 26 minutes. He was quick, steady, and looked ahead.

"One day an eagle went with the other birds to see who could fly the highest . . . (Next three sentences correct) . . . When he came back the others were waiting for him; and when he touched the ground a linnet flew off his back where the thief had hidden and said that he himself was the strongest bird. "I am stronger than you are," said the linnet, "for not alone did I fly as high, but as he began flying downward, then I left my hiding place and flew up a little higher," etc.

The following is a sample of the work of a boy of the same age, one of Terman's "stupid" group, who worked for 25 minutes at the same text. Save in one or two easy sections, his 'completions' make no sense at all. He worked by phrases only.

"One with the eagle and with the small birds and see who could fly the highest, and agreed and he who will fly the highest should be called the strongest they All started in the same place and whent away among the clouds . . . After a while he decided that the king of the little bird and not only and he was so high, but he did the thing as well," etc.

General Results and Conclusions.—(1) Performance in the completion test, as would be expected, improves noticeably with *age*. The results of Ebbinghaus are shown in Table 93.

TABLE 93

Dependence of the Completion Test on Maturity (Ebbinghaus)

SCHOOL GRADE	AVERAGE NUMBER SYLLABLES	AVERAGE PER CENT OF ERRORS
Untertertia	69	10
Quarta I.	49	17
Quinta	46	26
Sexta	32	33

Wiersma, however, was able to differentiate the relative influence of mere age, or maturity, and degree of school training (*Entwickelung*), and he then found that, while some correlation existed between age and performance in this test, a much greater correlation could be demonstrated between school training and performance (of children of the same age, those do better who are in higher classes), so that the table just quoted from Ebbinghaus may be looked upon as, in the main, a demonstration of this latter correlation.

(2) According to Wiersma, *sex* differences can not be made out with certainty.

(3) *Practise* may, according to Wiersma, effect an improvement in efficiency in the completion test that may be easily discerned after the lapse of 10 days, and even after an interval of 6 weeks. As a consequence, it is evident that, in making use of this test for comparative work at different periods, steps must be taken to eliminate or compute the practise-effect.

(4) In Ebbinghaus' Breslau investigation, no *fatigue* effects could be made out as the result of the five-hour session in the case of the upper classes, or at least, if fatigue were present, it was over-

balanced by practise: in the lower classes (10–12 years), however, the presence of fatigue was demonstrated.

It seems evident that the texts employed by Ebbinghaus were too easy for some of the upper classes tested by him: the effect was, naturally, to obscure the influence of fatigue and other factors.

Wiersma compared performance before and after a 10-days vacation, but he expresses his belief that the marked improvement exhibited by the pupils at the second test was largely due to practise. It is unfortunate that proper measures have been not taken to eliminate the practise error in these, and in other applications of the completion test.

In the opinion of Kraepelin, the Ebbinghaus test is to be regarded more as a device for exploration than as a decisive and accurate device for measuring fatigue, for, in the first place, no systematic study has yet been made of the relations between mental fatigue and the complex activities concerned in this test, and secondly, the evaluation of the errors is so difficult and their scoring so arbitrary that the test is not well designed for single applications (*Stichprobe*) and statistical treatment.

(5) The correlation with *intelligence* is, according to Ebbinghaus, clearest in the lower classes, and progressively less clear as the higher school grades are reached. When the Breslau pupils were divided into three groups—best, average, and poorest—on the score of class standing, their average rank was, in terms of quantity of work, 56, 48, and 43, respectively, and in terms of quality of work (percentage of errors), 17.3, 20.8, and 26.3 per cent, respectively.

Wiersma confirmed the existence of a correlation between capacity in the completion test and native ability (*Begabung*), both in tests at a teachers' seminary (ages 14.5 to 19.5) and at a continuation school (ages 12 to 15).

The author, however, has found no correlation between ability in this test (Text No. 1) and the class standing in psychology of 45 college students.

Ebbinghaus believes that the correlation between the completion test and intellectual ability may become obscured (1) because the test puts a premium upon speed of work, whereas the school grade is based on work that permits of a slower pace; (2) because, in some part, standing in the test may depend upon purely formal linguistic skill or verbal dexterity—a form of ability which, he thinks, has but a limited scope in school work, and (3) because the text selected for the test may be too easy.

In the author's opinion, these reservations are scarcely in order, in so far as Ebbinghaus implies that school grades are inferior to his test as a measure of intellectual ability and asserts that linguistic readiness plays no part in the determination of school grades.

The author is inclined, rather, to agree with Terman when he says: "My experience with the test causes me to regard it favorably; but, like all others, if taken alone, it can give only a partial account of the subjects' ability. It certainly does indicate something as to the general command of language. I am inclined to think that somewhat mechanical activities like memory and association, as distinguished from synthetic or combinative processes, play a relatively more important rôle in this test than Ebbinghaus assigns to them. Indeed, verbal memory, in the broad sense, would seem to be the chief factor in success." Incidentally, ability to spell, degree of familiarity with the type of literature from which the selection is taken, and the way in which *S* happens to go about the test may all affect his rank. Indeed, it is possible that a very original *S*, one with a spark of literary invention, might fare relatively poorly.

(6) *Other correlations*, as determined by Krueger and Spearman, are as follows: completion test and pitch discrimination, 0.81; completion test and adding, 0.93; completion test and the hypothetical 'central-factor,' 0.97.[1] The completion test has no correlation with a test of memory span (*Auswendiglernen*). The extremely high correlation with the 'central-factor' is of special interest, since it demonstrates a very close dependence of performance in this test upon a certain hypothetical psychophysical capacity, presumably akin to plasticity of the central nervous system, which, in the opinion of these authors, is, for each individual, a fundamental conditioning factor in the performance of various forms of mental activity.

NOTES.—It is evident that the outcome of the completion test hinges largely upon the degree of difficulty of the text employed: too difficult or too easy texts are alike undesirable, for the former convert the test into a blind puzzle, while the latter fail to bring out characteristic individual differences of ability.

To use the test on an extensive scale, therefore, we need to have at hand a number of texts that have been standardized by comprehensive trials with groups of *S*'s of both sexes, various ages, and

[1] These correlations are 'pure' correlations, computed from 'raw' correlations by the use of corrective formulas such as have been described in Ch. III, pp. 41–44.

various degrees of capacity and training. In other words, we need a series of norms of performance, or 'coefficients of difficulty,' as it were, for an adequate number of prescribed texts.

The difficulty of making comparisons between the results of different texts applied at different times may be further reduced by always permitting each *S* to finish each text, and by distributing the texts to be compared, in such a manner as to eliminate by subsequent computation whatever error arises from this difference of material.

The Lipmann-Wertheimer modification of the completion method is essentially as follows: a test-story is read to *S* to supply him with certain information which he is supposed thereafter to conceal. He is subsequently given for completion a mutilated text, the elisions of which are so arranged as to trap him into introducing facts from the test-story which he is trying to conceal.

REFERENCES

(1) H. Ebbinghaus, Ueber eine neue Methode zur Prüfung geistiger Fähigkeiten und ihre Anwendung bei Schulkindern, in Z. P., 13: 1897, 401-459.

(2) T. Elsenhans, Nachtrag zu Ebbinghaus' 'Kombinationsmethode,' in Z. P., 13; April, 1897, 460-3.

(3) E. Kraepelin, Ueber Ermüdungsmessungen, in A. G. P., 1: 1903, 9-30, especially 17.

(4) F. Krueger and C. Spearman, Die Korrelation zwischen verschiedenen geistigen Leistungsfähigkeiten, in Z. P., 44: 1907, 50-114.

(5) O. Lipmann and M. Wertheimer, Tatbestandsdiagnostische Kombinations-versuche, in Zeits. f. angew. Psych., 1: 1907, 119-128.

(6) M. Meyer, An English equivalent of "Combinationsmethode," in Jour. Phil., Psych. and Sci. Meth., 9: 1909, 688.

(7) L. Terman, Genius and stupidity, in Pd. S., 13: Sept., 1906, 307-373, especially 342-7.

(8) E. Wiersma, Die Ebbinghaus'sche Combinationsmethode, in Z. P., 30: 1902, 196-222.

TEST 49

Interpretation of fables.—E. J. Swift has suggested that mental ability may be measured by determining the capacity to interpret the total situation given in a typical fable. Swift's

three test-fables have been also employed by Terman, in addition to 12 other fables, in his study of the comparative efficiency of bright and stupid children.[1]

MATERIAL.—The following are the three fables used by both Swift and Terman.

Fable 1.

THE BOY AND THE FILBERTS.

"A boy once thrust his hand into a pitcher nearly filled with filberts. He grasped as many as his hand could possibly hold; but when he tried to draw out his closed fist, the narrowness of the neck prevented him from doing so. Unwilling to lose his nuts, yet unable to get them by drawing out his hand, he burst into tears and bitterly lamented his hard fortune."

Fable 2.

THE HORSE AND THE RIDER.

"A cavalry officer took the greatest of pains with his horse. As long as the war lasted, the horse was looked upon as a companion and fellow-helper. He was carefully groomed every day and fed with oats.

But when the war was over, the allowance of grain and hay ceased, and the horse was fed with chaff, and whatever he might find by the wayside. He was made a drudge, too, and often forced to carry loads much too heavy for his strength.

When, in course of time, war was again proclaimed, the soldier brought his military trappings, and put them on his horse; and, after putting on his own coat of mail, he mounted to ride to battle.

But the horse, no longer equal to the burden, fell down straightway under the weight."

Fable 3.

THE FISHES AND THE PIKE.

"The fishes in a pond brought an accusation against the pike who were eating them up. The judge, an old pike, said that their complaint was well founded, and that in the future, to make things right, he would allow two ordinary fish every year to become pike."

[1] The details of these 12 fables are not published, but the reader may form an idea of their nature by consulting Terman's original article. It would seem desirable that further tests should be made in order that a series of test-fables of varying difficulty might be standardized by application to large groups of normal children of various ages.

Method.—Read the first fable to *S*. Ask him then to tell what he thinks of the boy, and why he thinks so. Use a few discreet questions, if necessary, to be positive as to his interpretation of the story. Record his statements verbatim. Continue the test by reading Fable No. 2, and asking *S* what he thinks of the cavalry officer and why. Similarly, ask his opinion of the plan of "Judge Pike."

Variations of Method.—So far as the mechanics of method are concerned, this test may be conducted with groups of *S*'s by using written answers, but it appears evident from the results that this method is to be avoided whenever possible.

Treatment of Data.—*S*'s standing must be determined by observation of the readiness with which he offers an interpretation to each fable, and by the soundness and excellence of this interpretation. For quantitative comparison, *E* will find it advisable to assign the rank 5 for a perfectly satisfactory answer, 1 for a complete failure to 'see the point,' and intermediate grades in proportion. For a final score, add the rank obtained by *S* in each of the three trials.

Typical Results.—Terman quotes the following 'interpretations' of Fable 3. Sample of a 'good' answer (Grade 5): "Not fair." (Asked why) "Because the more pike there became, the faster they would eat the little fish." Sample of a very poor answer (Grade 1): "It might be better for the little fish in some ways—they would not be eaten up." Sample of answer of medium quality (Grades): "That was all right, but then if only two are changed to pike, the others would get eaten up."

Results.—(1) *Correlation with school proficiency.* Swift reports that the fable-test showed no superiority for the 'bright' children; that, on the contrary, in Fable 3, which, he says, requires "a distinctly intellectual process," the dull group excelled the bright group, while the answers of boys in an industrial (reform) school "distinctly outranked those from both of the other groups, but especially the ones from the 'bright' division, in the penetration and versatility that they showed."

Terman's tests showed, on the contrary, that his 'dull' group was distinctly inferior to his 'bright' group, for, as he says: "in the first place, they more frequently miss the point of the story

altogether,'' and "in the second place, the dull boys are plainly deficient in degree of abstraction. Even when they give an approximately correct interpretation, they usually express it in the concrete terms of the given situation, instead of generalizing it," *e.g.*, the following answer for Fable 2: "It teaches a man when he has a horse to keep it and use it well."

This lack of accordance is attributed by Terman to some fault in method on the part of Swift. "I should judge," he says, "that his results would have been different if he had been able to take his cases individually, instead of collectively."

By way of illustration, the following quantitative results may be quoted. Swift reports that, with Fable 3, 27% of his 'bright' group, as contrasted with 9% of his 'dull' group, thought the plan wise and just. Again, only 15% of the bright children, as contrasted with 30% of the dull children, pointed out that the plan would not help the fishes that were not turned into pike. When scored according to Terman's plan,[1] his bright *S*'s averaged for Fables 1, 2 and 3, the ranks 1.3, 3.17, and 2.83, respectively, while his dull *S*'s averaged the ranks 1.86, 4.57, and 4.57 for the same three fables.

(2) *Dependence on age.* Terman suggests that "what is tested by the interpretation of fables is, in part at least, that general change of mental horizon that comes with increased experience and dawning maturity." In other words, the outcome of the test is conditioned, as one might have expected, by age and experience, as well as by native ability.

(3) *Types of replies.* Swift states that "the answers from the public school children lacked individuality: they were conventional, while those from the reform school gave evidence of spontaneity and resourcefulness," and he draws from this the dubious conclusion that "the question may be seriously raised whether the schools do not train children to stupidity."

REFERENCES

(1) E. J. Swift, Standards of efficiency in school and in life, in Pd. S., 10: 1903, 3-22, especially 3-6.

(2) L. M. Terman, Genius and stupidity: a study of some of the intellectual processes of seven 'bright' and seven 'stupid' boys, in Pd. S., 13: 1906, 307-373.

---

[1] Terman gave the score 1 for a satisfactory answer, 5 for a complete failure.

CHAPTER XII

Tests of Intellectual Equipment

The tests of this chapter differ from other mental tests described in the present volume in that they measure, not the efficiency with which certain typical mental activities or mental processes can function, but rather the number of ideas that an individual possesses. In other words, their purpose is not to measure what the individual can do, or how well he can do it, but what he knows about—to take a census, as it were, of his stock of information. G. Stanley Hall's study of the content of children's minds on entering school[1] is, perhaps, most nearly allied in type and conception with the tests which are here presented.

The first test is designed to secure an estimate of the number of words in the reading vocabulary of the individual tested, the second to secure an estimate of the number of subjects (disciplines, phases of human activity) with which the individual has an exact or an approximate acquaintance.

TEST 50

Size of vocabulary.—Since nearly all thought and expression is couched in linguistic form, and since the intellectual progress of the child at school is, in a sense, a process of augmentation of his vocabulary and of refinement in its use, it seems not unreasonable to assume that the determination of the size of this vocabulary will be of significance and value in estimating his general intellectual status.

Experiments conducted by Kirkpatrick have shown that an approximate determination of what might be termed the *vocabulary-index* can be secured by the use of the relatively short and

[1]See his *Aspects of child life and education*, Boston, 1907.

simple test that is described herewith. By extending the tests, the usual comparative study may be made, and the index may be related to its conditioning factors—age, sex, school standing, extent of reading, general ability, etc.

This test has, up to the present date, been applied only by the originator, Professor Kirkpatrick, and by the author, though with some variations of method.

MATERIAL.—Printed list of 100 words, as employed by Kirkpatrick and by Whipple.

PRELIMINARIES.—In accordance with Kirkpatrick's plan, several preliminary exercises are employed, in order, on the one hand, to obtain data with regard to *S*'s general familiarity with words, his range of reading, etc., and on the other hand, to instill in him an attitude of caution in undertaking the vocabulary-test proper. These preliminary exercises are as follows:[1]

(1) Ask *S* to write the opposite of the following terms: *good, long, break, rude, simple, permanent, particular, permit, obnoxious, genuine.*

(2) Ask *S* to tell (orally or in writing) what the following words mean: *abductor, baron, channel, decemvirate, eschar, amalgamation, bottle-holder, concatenate, disentomb, filiform, gourd, intercede, matting, page, hodman, lanuginose, muff, photograph, scroll, tycoon.* (Where words have more than one meaning, all are to be given.)

(3) Secure from *S* a list of all the papers and magazines that he is in the habit of reading.

(4) Secure from *S* the names of the books that he has read during the past 6 months.

(5) Ask *S* which of these books he liked best, why he liked it, and to give some account of what it was about.

(6) Ascertain the birthplace of *S*'s parents, his school grade, and his favorite school subjects.

METHOD.—Hand to *S* the printed test-slip: ask him to read the instructions and to mark the words carefully in accordance with them.

TREATMENT OF DATA.—The number of 'plus' marks indicates the vocabulary-index in terms of per cent of words known. The

[1] To follow the plan, the first two exercises, at least, should be given whenever grade pupils are tested, and all five if time permits.

absolute vocabulary is then computed, following Kirkpatrick, by assuming the total number of words in the language (Webster's Abridged Dictionary) as 28,000.

VARIATIONS OF METHOD.—To study *S*'s tendency to overestimate or underestimate his vocabulary, *E* may follow the plan used by the author with college students, of giving the check-definition test after the vocabulary-test.

(1) Give *S* the vocabulary test without suggesting that he may be called upon to justify his marking.

(2) When the marking has been completed, and the slip is in *E*'s hands, submit to *S* the following list of words, with a request that each word be defined. Allow 20 min. for written definitions.

Definition-list.[1]

abductor	interdict	amalgamation	lanugo
abet	interim	amanuensis	lanyard
baroscope	mattock	amaranth	mufti
chanticleer	maturate	bottomry	photo-lithograph
chaos	pudgy	concatenate	rejoinder
decemvirate	scruff	disentrance	skysail
eschar	scrunch	disepalous	tendinous
escheat	subcutaneous	disestablish	tendril
eschalot	tycoon	filiform	virago
gourd	tymbal	hoecake	virescent

TREATMENT OF DATA.—For each *S*, ascertain from the definition-test: (1) the number of words not defined, (2) the number of words wrongly defined. (3) Add these to find the total number of words unknown in the list of 40. (4) Consult the vocabulary test-slip to see whether any words outside the list of 40 are marked unknown.[2] (5) Consult the vocabulary-slip again to see whether any words thereon are marked doubtful and have not been cleared up by the definition-test: consider these as unknown. (6) Add all the unknown terms to determine the final corrected vocabu-

[1] *Concatenation, lanuginose* and *lanuginous,* of the vocabulary-test, can, of course, be checked off by the definitions given for *concatenate* and *lanugo.*

[2] In a test of Sophomores and Juniors in college, we were surprised to find the following words in this category: *barque, barouche, boudoir, disentomb, filigree, hodman, pagoda, rejuvenate, scroll, sub-let, tenderloin.* These words, then, it seems, would have to be added to the 40 to secure a comprehensive list of possibly unknown words.

lary-index. (7) Compare this index with the index indicated by S on the vocabulary-slip to see whether S has over- or underestimated his vocabulary, and to what degree.

RESULTS.—(1) Kirkpatrick's computation of the *average vocabulary* is shown in Table 94. The author's results, based on 70 college students (16 men and 54 women), aged 16 to 25 years, indicate an average vocabulary of 21,728 as based on the uncorrected estimates of the students, and of 20,512 as based on the corrections supplied by the supplementary definition-test.[1]

TABLE 94

Average Vocabulary in Relation to Scholastic Status (Kirkpatrick)

SCHOLASTIC STATUS	VOCABULARY	SCHOLASTIC STATUS	VOCABULARY
Grade II	4480	Grade IX	13,400
Grade III	6620	High school, 1st year	15,640
Grade IV	7020	High school, 2d year	16,020
Grade V	7860	High school, 3d year	17,600
Grade VI	8700	High school, 4th year	18,720
Grade VII	10,660	Normal-school pupils	19,000
Grade VIII	12,000	College students	20,120

(2) In the author's *definition-test*, no word of the 40 was correctly defined by every student, and since, as has been noted, there remained 16 other words that were unknown or doubtful, it follows that only 44 of the 100 words in Kirkpatrick's list were certainly known by every one of 70 college students.

(3) There is wide *individual variation* in the size of the vocabularies of students of the same age and scholastic status. This variation is shown by the distribution in Table 95. The largest college-student vocabulary found by the writer is 24,920 (89 per cent); the smallest is 16,240 (58 per cent), or approximately the vocabulary assigned by Kirkpatrick to the average 2d-year high-school pupil.

(4) No positive *sex differences* have been established, though

[1] For the size of the vocabularies of children up to 4 years of age, consult Whipple (5), where there will be found a summary of the results of numerous investigations.

TABLE 95

Distribution of Corrected Vocabulary-Index. 70 College Students (Whipple)

INDEX	55–59	60–64	65–69	70–74	75–79	80–84	85–89
No. of cases	1	6	13	22	19	5	5

Highest index, 89%. Average index, 73.26%. Lowest index, 58%.

there is a suggestion of superiority of boys over girls, and of men over women.

(5) In general, pupils that read the most books and magazines have the largest vocabularies.

(6) Kirkpatrick found a tendency toward positive *correlation* between class standing (teachers' grades) and vocabulary-index: "those ranking high in scholarship knew on an average about 5 per cent more words than those ranking low in scholarship." The author found a more decided correlation ($r = +0.45$, P. E. $= 0.06$) between the index of 58 college students and their grades in his classes in educational psychology.

(7) When no precautionary measures are taken to offset the tendency, the determination of the vocabulary-index is commonly affected by *overestimation*. Inspection of Table 96 will show that 59 of the 70 college students examined by the author overestimated, while but 10 underestimated their vocabulary: the largest overestimation was 18 per cent; the largest underestimation was 4 per cent. Since 20, or more than ¼ of the students overestimated

TABLE 96

Overestimation of the Vocabulary-Index. 70 College Students (Whipple)

PER CENT OVERESTIMATED	NUMBER	PER CENT OVERESTIMATED	NUMBER
18	1	5	7
15	2	4	6
14	3	3	7
13	1	2	10
12	3	1	9
11	1	0	1
9	2	—1	6
8	2	—2	2
7	1	—4	2
6	4		

by 5 per cent or more, it is evident that, without a somewhat elaborate definition check, the reliability of the vocabulary-test is distinctly lessened.[1]

(8) The definition-test reveals an unexpectedly large number of *erroneous definitions.* The source of these errors may frequently be traced to confusion with words of similar appearance or to fancied etymological derivations. The following list shows typical errors in definition by college students: the assumed source of confusion is indicated by the terms in parentheses after the definitions:

amanuensis—poet laureate, lovingness (amativeness).
amaranth—a precious stone (amethyst).
abet—although (albeit), a wager (a+bet), diminish (abate).
bottomry—the art of bottoming chairs, deceit, bottom of anything.
chanticleer—one who sings a loud song, one who leads a chant.
decemvirate—composed of five, count out by tens, formerly a group of ten men, but any number now.
disentrance—failure to enter.
disepalous—apart from the head, without shoulders.
gourd—reward (guerdon), to slash or whip (goad), morning glory.
interim—time between two reigns (interregnum).
lanugo—a kind of language.
lanyard—yard where leather is tanned (tanyard), yard about the lane.
mattock—a lock of hair (matted locks?), a kind of bird, a sort of rug, a kind of robe (cassock).
maturate—to ripen (mature), to matriculate.
sky-sail—a sail in the sky, a kite.
tycoon—a violent wind (typhoon), an animal, a silk-worm, a natural phenomena (sic).
tendril—a membrane connecting two bones (tendon).
tendinous—capable of endurance (tenacious?).
scrunch—a good for nothing person (scrug?).
virago—a kind of bird (!) (vireo), a disease, giddiness (vertigo).
virescent—sparkling (iridescent), of or pertaining to man (!) (virile).

NOTES.—(1) The greatest source of unreliability in the vocabulary test lies in individual differences in the subjective standard

[1] This result may be compared with Kirkpatrick's conclusion that very young children are apt to underestimate because the isolated words of the list fail to arouse associations such as they would if they had a context. Again, when Kirkpatrick defined the words of the list to normal-school students, he found that the errors of over- and underestimation tended to cancel one another; while, when college classes defined 20 words, 114 of 246 students (about 46 per cent) correctly defined the same proportion that they had marked as known, and only 7 per cent erred by as much as 3 in 20.

employed by different *S*'s in marking their lists: some *S*'s mark, as known, words which are little more than familiar; others mark words as known only when they can define them accurately. Kirkpatrick suggests that, to minimize this source of error, the directions might be amplified as follows: "Count as known all words whose meaning would be known, without consulting a dictionary, when read in a sentence."

(2) This leads one to say again that, especially in the case of young children, there may be a tendency toward underestimation of the vocabulary because isolated words sometimes fail to arouse the interpretative meanings that they would arouse at once in their customary context. In so far as appeal to the ear as well as to the eye is of assistance to young children whose vocabulary is largely auditory, this source of error might be partially offset by reading the list aloud to them.

(3) In grading the definition-test, it is at times rather difficult to decide from the definitions whether *S* does, or does not know the meaning of a word with sufficient exactness to be credited with knowledge of the term in question. In general, it is better, in consideration of the difficulty of accurate definition and of the short time usually available for this part of the test, to err on the side of leniency.

Thus the following definitions might be accepted: 'disestablish—to overthrow,' 'decemvirate—a body of ten,' 'mattock—a garden tool,' 'amaranth—a flower;' while the following ought, in our opinion, to be disallowed: 'lanyard—one of the spars of a ship,' 'decemvirate—Roman civil officer,' 'gourd—a hollow vessel from which to eat and drink,' 'concatenate—to argue,' 'baroscope—an instrument for measuring something.'

REFERENCES

(1) E. H. Babbitt, A vocabulary test, in Pop. Sci. Mo., 70: 1907, 378.

(2) E. Kirkpatrick, Number of words in an ordinary vocabulary, in Science, 18: 1891, 107–8.

(3) E. Kirkpatrick, A vocabulary test, in Pop. Sci. Mo., 70: 1907, 157–164.

(4) G. M. Whipple, Vocabulary and word-building tests, in P. R., 15: 1908, 94–105.

(5) G. M. Whipple (and Mrs. Whipple), The vocabulary of a three-year-old boy, with some interpretative comments, in Pd. S., 16: 1909, 1–22. (Contains references to 27 articles on children's vocabularies.)

TEST 51

Range of information.—The words that comprise Kirkpatrick's vocabulary test are intentionally selected by chance: some of them, like *page,* happen to be most ordinary and everyday terms; others, like *lanuginose,* are unusual, technical terms. The extent of *S*'s acquaintance with words of the latter kind depends almost entirely upon the nature of his school training, or upon the quantity and type of his general reading.

The range of information test has been devised by the author as an extension of the vocabulary test. The hundred test-words have been selected, not by chance, but by careful consideration, and in such a manner that each shall be representative of some specific field of knowledge or activity, in the sense that if *S* has made himself familiar with a given field, he will almost certainly know the word selected from that field, whereas if he has not made himself familiar with the field, he will almost certainly not know the term, or at least will not have such knowledge of it as to enable him to define it exactly. Thus, general knowledge of American history is tested by the name *Anthony Wayne,* knowledge of French by *aujourd'hui,* of chemistry by *chlorine,* of ethics by *hedonism,* of golf by *midiron,* of social usages by *R. S. V. P.,* of the technique of photography by *f-64,* etc.

MATERIAL.—Specially prepared test-blank containing 100 test-words, directions for marking them, and a request for 10 definitions.

METHOD.—Place the blank in *S*'s hands: ask him to read the directions through twice before marking the words, and call his attention to the request for definitions as printed below the test-words. Let him take his own time.

For exact results, *S* should afterward be required to define every word that he has marked *D,* and to explain or attempt to explain every word that he has marked *E* or *F.* This check test should, by preference, be conducted orally. In practise, however, especially when testing by the group method, such careful checking may prove too onerous: erroneous definitions may then be neglected, or the quantitative data may be revised by discounting on the basis of the percentage of error revealed in the definitions actually

given. Or, again, E may, after the test is concluded, define the 100 words, and let each S revise his own paper by placing a second series of marks *after* each word to indicate the manner in which he should have marked it. A comparison of the sums of the D's, E's, F's and N's of the first, and of the second series will then show approximately the extent and nature of the error due to ignorance or misunderstanding of the real meanings.

RESULTS.—(1) Typical quantitative results, as obtained by the author from some hundred cases, are embodied in Table 97. Here it is evident that advance in school training, together, of course, with increased maturity, is paralleled by an increase in the number of technical terms that can be defined (F), explained (E), or that are at least familiar (F), and by, of course, a corresponding decrease in the number of terms that are new or unknown (N).

TABLE 97

Dependence of Range of Information on Academic Status (Whipple)

ACADEMIC STATUS	NUMBER	D	E	F	N
Graduates	4	39.0	21.0	12.2	27.0
Seniors	5	20.6	17.2	25.2	37.0
Juniors	10	24.8	12.0	23.7	39.5
Sophomores	30	17.7	12.7	17.3	52.2
High School	52	6.8	7.6	16.3	69.3

(2) *Comparison of the sexes* indicates the superiority of the range of information of men and boys over women and girls, as is indicated in Table 98.

TABLE 98

Dependence of Range of Information on Sex (Whipple)

	NUMBER	D	E	F	N
Men	44	15.79	11.98	18.22	54.02
Women	57	12.21	9.42	17.19	61.17

(3) The results just figured are 'raw' results: strictly speaking, these should be revised on the basis of an extended series of definitions, as recommended in the Vocabulary Test, since an inspection

of the definitions and explanations actually given reveals in the majority of the papers one or more errors, due in the main to confusion with words of similar appearance or to fancied etymological derivations. The following list shows typical errors in definition by college and high-school students: the assumed source of confusion is indicated by the terms in parentheses after the definitions:

ageratum—an aggregation of objects: the aggragate (sic) amount.

annealed—pressed or rolled out thin: molded together.

Anthony Wayne—a historic character who was hung in the cause of freedom for the blacks: a man who fought in the Revolution on the English side.

Babcock test—a device to ascertain whether or not cattle have tuberculosis.

base-hit—when the ball is hit and strikes a base or is caught there: a ball batted over a base: when the striker bats the ball into the pitcher's hands.

Bokhara—name of a place in Austria.

cantilever—a bar with a hook in one end by which lumbermen roll logs (canthook).

catalepsy—a form of disorder of the nervous system which causes fits or convulsions (epilepsy). (Similar statements given by 15 persons.)

chamfer—the tree from which camphor gum is obtained: this is the simplified spelling of it (!). (The confusion with camphor was found in 4 papers.)

clearing-house—a sale that takes place when a store wishes to dispose of its stock (clearing sale): a place where clearing papers are given to vessels to enable them to leave the harbor (customs house+clearing of vessels): picking up everything to move; taking everything out of the house: a place used by express companies to sell uncalled-for goods: a house where goods are made ready to be delivered.

cotangent—name of one of two tangents drawn to a circle from the same point without the circle: one lying alongside of (contingent): straight line drawn to touch a circle at one point (tangent).

dibble—to get just a smattering of some subject, as to dibble in medicine or politics (dabble): to do with divided interest (dawdle).

dryad—a priest of early English times (druid).

entrée—first course at a banquet, usually soup: something in the way of food, new and out of season: when the waiter brings in a new course it is called an entrée: French for 'to-day': French for 'between' (*entre*).

Eocene—the term applied to one of the early ages of civilization.

Euclid—a book written by Vergil (Æneid): name given to certain trees (eucalyptus): an ancient Egyptian who studied geometry: name of an avenue in Cleveland, Ohio.

f-64—means the temperature is 64 degrees above zero, Fahrenheit.

f. o. b.—cash on delivery (c. o. d.): forward on board.

hydraulic press—a kind of air-pump, rather complicated, operated by suction and pressure: a machine for washing dirt from gold or from steep slopes (hydraulic mining): the force with which water flows upon or against a thing, as a paddle wheel.

impressionism—when a man imitates the looks or actions of another: the art of exciting an impression.

infusoria—a chemical herb (infusion?).

kilogram—the greatest quantity in the metric system: French measure of distance (kilometer): French unit of liquid measure: the weight of a cube of water whose dimensions are a kilometer.

Les Misérables—a French tragedy written about the last part of the 17th century by Racine, one of the famous French writers: French work written by George Sand, author of Le Diable.

linotype—the product of a certain method of making prints from photographs.

Millet—a blind poet (Milton).

natural selection—in nature each animal selects its mate, a device for building up a stronger race.

ohm—German word for uncle (*Oheim*).

Polonius—a prominent character in Julius Cæsar.

pomology—the study of the palm of the hand, used by fortune tellers (palmistry).

tort—French word for ugly (*tors?*).

triple expansion—the expanding of anything three times its normal size.

Utopia—a silk factory.

way-bill—a bill that is being considered.

Zionism—same as Dowieism.

(4) A comparison of scores made by 18 summer session students, before and after the definition by E of the 100 terms, shows the following averages: first marking, D = 20.39, E = 14.77, F = 18.39, N = 46.44; second marking, D = 19.77, E = 20.22, F = 19.55, N = 40.44. So far as these S's are concerned, then, it appears that at first they had overestimated terms definable and, more particularly, terms unknown. The principal effect of E's explanations was to increase by about 6 per cent the number of terms marked as explainable, and to decrease by 6 per cent the number of terms marked as unknown.

REFERENCE

G. M. Whipple, A range of information test, in P. R., 16: 1909, 347-351.

CHAPTER XIII

Serial Graded Tests for Developmental Diagnosis

The tests which have been described in the preceding chapters are mainly of an isolated character, in the sense that each test has been devised independently, and they are designed primarily for the examination of adults or normal children. The tests which are described in the present chapter are, on the contrary, designed primarily to determine the mental insufficiency or retardation of backward and abnormal children, and they are arranged systematically, in a graded series, such that each test or group of tests represents an advance in difficulty over its predecessor, while the series as a whole corresponds to progressive stages or degrees of mental development.

There is no doubt but that, as Meumann contends,[1] this idea of a progressive series of tests is of significance for the investigation of normal as well as of abnormal children, and, in fact, Binet and Simon have not limited the use of their series of tests to abnormal children, but have sought to establish standards of performance for normal children as well.[2]

TEST 52

De Sanctis' graded series.—Sante de Sanctis, of the University of Rome, has sought to arrange a graded series of tests, which shall be at once simple and comprehensive, applicable to

[1] E. Meumann, *Vorlesungen zur Einfuhrung in die exp. Padagogik*, i., p. 391.

[2] There is, of course, no particular reason why the tests mentioned in this chapter should be the only tests arranged in a graded series. What is really needed, as we have already insisted, is that all of the tests that are described in this book, in so far as they prove of permanent value, should, by the collaboration of investigators, be so standardized that we may know the normal outcome of every test for children of all ages, just as we already know the norms for all ages, and the distribution by percentile grades, in most of the anthropometric tests.

any feeble-minded child from 7 to 16 years old, and so selected as to throw into relief the differences in the intellectual capacity of mentally defective children.

MATERIALS.[1]—Five wooden balls, 50 mm. in diameter, painted red, orange, yellow, blue, and green, respectively. Three wooden pyramids. Two wooden parallelopipeds. A set of 12 cubes

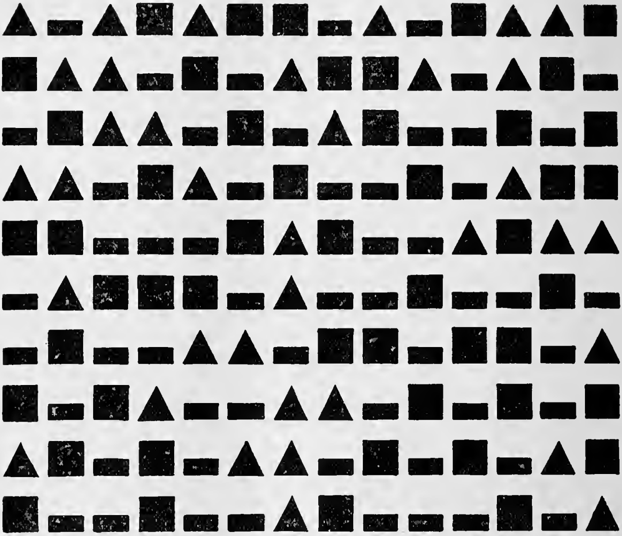


FIG. 58. DE SANCTIS' FORM-TEST.
About $\frac{1}{4}$ actual size.

ranging in size from 10 to 80 mm. A small black cube. Test-card of forms (Fig. 58). Stop-watch. Cardboard screen (or cloth cover).

[1] It is unfortunate that de Sanctis gives little more than a sketchy description of his materials and method. We have, therefore, followed Goddard in planning the material and in arranging the details of procedure.

METHOD.—*General directions.* See that the child is comfortable and happy, not fatigued, frightened, or ill. Make a pleasant game of the whole procedure.

Introduce a pause of 40 sec. between No. 1 and No. 2, and a pause of 1 min. between the other tests.

In the effort to catch *S*'s attention and to explain the tests to him, instructions and questions may be repeated till given thrice: if *S* then fails, he is considered to be incapable of passing the test—save that, if the failure seems to be due to a temporarily unfavorable disposition, the entire series may be repeated several days later.

No. 1.

Put the five balls on a table. Say to *S:* "Give me a ball." Record with the stop watch the time taken by him in responding. Put the ball away at once, with the others, behind the screen.

No. 2.

Show the five balls again, and ask *S:* "Which ball did you give me"? Record the time of response.

No. 3.

Put on the table, screened from *S*, the 5 cubes of the same size, mixed in with the 3 pyramids and the 2 parallelopipeds. Show *S* another cube (the one of the series of 12 which is nearest in size to the 5 on the table), saying: "Do you see this bit of wood? Well, pick out all the bits here on this table that look like it." Remove the screen, and time *S*'s performance.

No. 4.

Show the test-card of forms and the small black cube, saying: "Look at this bit of wood; is there anything on this card that looks like it"? If *S* answers correctly, direct him: "Pick out all the things that look like it." (Or: "Pick out all the squares on the card.") *S* may mark the squares with a pencil, or, preferably, indicate them with a small pointer. Record the time and the number of errors.

No. 5.

Put on the table, at different distances from *S*, the series of 12 cubes of varying sizes. *S* is told: "Here are a lot of bits of wood that look like the things you were just picking out on the card. Look at them carefully, and tell me how many there are, which is the biggest, and which is farthest away from you."[1] Take the time: record errors.

No. 6.

With all the test-material concealed from view, ask *S* these four questions, but allow time to answer each one before the next is asked.

(*a*) "Are big things heavier or lighter than small things"?

(*b*) "How does it happen that small things are sometimes heavier than big things"?

(*c*) "Which looks bigger, a thing that is close by, or a thing that is far away"?

(*d*) "When things are far away, do they just look smaller or are they really smaller"?

Measure the times of response, and record the answers verbatim. In the 3d and 4th questions, *E* may assist young children with no training by illustrating with the objects selected by them (in No. 5) as the largest and the most distant.

Interpretation of Data.—De Sanctis determines the grade of development of children as follows; (*a*) Ability to pass only Nos. 1 and 2 indicates intellectual defect of a high degree. (*b*) Ability to pass Nos. 1 to 4 only, or No. 5 with difficulty and with many errors, indicates a moderate degree of defect. (*c*) Ability to pass No. 5, but not No. 6, indicates but a slight degree of "mental insufficiency." (*d*) Ability to pass all 6 tests in normal time indicates normal intelligence.[2] A child may pass all 6 tests and

[1] This test is not carefully described. Apparently, de Sanctis uses a distance not over 2 cm. between the farthest and the nearest cube, and asks the three questions together, in a single sentence, as printed. Goddard arranges the cubes so that all shall be within easy reach of the child, but that one shall be distinctly farther than the rest. Goddard also, apparently, makes three separate questions of those that de Sanctis joins together. The difficulty of the test will evidently be considerably affected by these changes in arrangements of the material, and in the form of procedure.

[2] No results have been published to show what times are to be considered normal. *E* must work these out by trial with a group of normal children.

yet be rated as 'backward' pedagogically; such a child may exhibit an abnormal character (flighty, undisciplined, incorrigible, etc.), but he is not 'backward' in the strict medico-psychological sense of the term.

RESULTS.—No statistical results have been published. De Sanctis and Decroly and Degand report that the tests give excellent results when applied to children from 7 to 16 years of age, and that these results accord closely with those of clinical examinations and with the pedagogical 'histories' of the children.

NOTES.—De Sanctis believes that by these tests one discovers directly (*a*) capacity of adaptation to situations, (*b*) memory for color, (*c*) capacity to recognize forms, and particularly to recognize the relation between a plane and a solid figure, (*d*) capacity to give persistent attention, (*e*) ability to enumerate objects and to judge their number, size, and distance, (*f*) ability to reason out relations between objects and their qualities, when they are not present to the senses, and (*g*) quickness of mental processes, particularly of perception, action, and thinking.

A certain amount of disturbance is undoubtedly introduced in the attempt to infer mental condition when the tests are applied to children of different ages, but this difficulty ought to disappear after E has had due experience in administering the tests.

Goddard thinks, too, that the effect of training is *apparently* not entirely eliminated: a child with rather good training *may* pass No. 6, and still be feeble-minded. However, if all the tests are passed with normal rapidity, the child may probably go back eventually into the regular school grades.

REFERENCES

(1) S. de Sanctis, Types et degrés d'insuffisance mentale, in A. P., 12: 1905 (1906), 70-83.

(2) H. H. Goddard, The grading of backward children, in The Training School, November, 1908. Also issued as a reprint from the New Jersey Training School, Department of Psychological Research, August, 1909.

TEST 53

Binet-Simon graded tests: 1905 series.—The object of this series is to provide a quick means for the psychological diag-

nosis of the grade of intelligence of a backward or abnormal child by means of 30 tests of a simple, but precise character, sufficiently varied in type to explore all the important phases of intellectual capacity (with special reference to judgment—good sense, initiative, adaptability), and of such a kind as to permit an intelligent investigator to form an independent estimate of the child's mental equipment. The tests are designed to measure native ability rather than erudition or scholastic attainment. They are to be administered individually, with suitable precautions to insure the goodwill and active coöperation of the child, and to avoid restraint or timidity. The tests are first applied to selected normal children of from 3 to 11 years of age and the insufficiency or retardation of abnormal or backward children is later estimated by comparison of their results with those of the normal children.

MATERIALS.—For No. 1: matches.

For Nos. 2–5: a small piece of milk chocolate, a small piece of white wood of similar dimensions, and a piece of paper in which the chocolate may be wrapped.

For No. 7: three familiar objects, *e.g.*, a piece of string, a cup, a key.

For Nos. 8 and 9: a set of 8 colored pictures, representing familiar scenes. [These lithographs are reproductions of illustrations in the "Jingleman Jack" book, and have been selected by Goddard as being specially adapted for the testing of defective children.]

For Nos. 10 and 13: two sheets of heavy white paper, each 15×20 cm; on one sheet two straight black lines, about 0.5 mm. wide, one 40, one 30 mm. long, drawn end on end, 5 mm. apart; on the other, two lines, each 40 mm. long, similarly placed.

For Nos. 12, 22, and 23: five cubical boxes, about 23 mm. in size and all of the same color, but loaded with shot so that they weigh 3, 6, 9, 12 and 15 grams, respectively. They are marked inconspicuously, so that their weights are not known to *S*.

For No. 17: two sets each consisting of 12 pictures[1] of familiar objects, mounted on a single sheet of cardboard.

[1] Binet and Simon used 13 pictures, Decroly and Degand three groups of 8 pictures each.

For No. 18: a card on which is printed the two designs shown in Fig. 59.

For No. 21: five sheets of heavy paper or cardboard, each 15× 20 cm. In the exact middle of each one are drawn in black ink two straight lines, 5 mm. apart, end on end; the one line is 30 mm., the other is 31, 32, 33, 34 and 35 mm. on each one of the five sheets respectively. Three sheets, each 20 × 30 cm. on which are drawn similarly a standard line 100 mm. and comparison lines of 101, 102, and 103 mm. respectively.[1]

For No. 29: two sheets of plain paper (say 20 × 26 cm.), scissors, pencil.

General Purpose, Method, and Interpretation of the 30 Tests.[2]

No. 1. Visual coördination.

Move a lighted match slowly before *S*'s eyes. Note whether he follows the movement with properly coördinated movements of the head and eyes. (As an accessory test, *E* may ring a bell behind *S* to see whether he adopts a listening attitude.)

No. 2. Prehension provoked tactually.

Place the small wooden cube in contact with the palm or back of *S*'s hand to see whether he can execute properly coördinated movements of grasping. See whether he can carry the object to his mouth.[3] Verbal directions may be used, *e.g.*, "Here is something for you: take it," etc.

[1] Binet and Simon used 15 sheets for the former and 12 for the latter: it is quite as simple to use a single sheet and reverse the place of the longer line without *S*'s knowledge. It is not clear in what manner the originators conducted this test.

[2] The order of presentation need not follow that here given; thus, No. 6 should really come first; No. 13 is divided between several other tests; the simpler ones may often be omitted; others may be shifted at convenience. The order here given corresponds with that assigned by Binet and Simon, and represents approximately the progressive difficulty of the tests.

[3] In our opinion, it would be better to avoid this reference to the mouth, since No. 4 hinges upon *S*'s avoidance of this movement, and it is perfectly easy to test the execution of the movement in the food test.

No. 3. Prehension provoked visually.

Repeat No. 2, with the object placed within *S*'s reach, but not in actual contact with his skin. This test may be combined with the next.

No. 4. Cognizance of food.

Show *S* successively the small bit of chocolate[1] and the piece of wood of similar dimensions. See which object he takes and tries to eat. If *S* turns away, refuses to try or makes defensive gestures, this usually indicates a very low degree of intelligence, unless his conduct is due to excessive timidity or excitement.

No. 5. Seeking food when a slight difficulty is interposed.

This test is designed to show the presence of rudimentary memory, volitional effort, and ability to execute simple coordinated finger movements. When sure that *S* knows the chocolate and desires it to eat, wrap up a small bit of it in a piece of paper; let *S* see this done, then give him the packet and note his action. Does he, for instance, throw it away, eat it whole, bite off the paper, pass it to someone else to unwrap, make a single futile trial, succeed quickly, or what?

No. 6. Execution of simple orders and imitation of gestures.

When *S* first enters the room, *E* should greet him with somewhat exaggerated politeness, extend the hand and say "Good-morning," to see whether *S* understands the gesture and responds to it readily. Ask *S* to be seated: drop something and ask *S* (by gestures as much as by words) to pick it up and hand it to you: try such simple commands as "stand up," "close the door," "come here," etc. For further tests, get *S*'s attention, then ask him to "Do as I do." Try clapping the hands, putting the hands on the shoulders, behind the back, rising on the toes, etc. This test should be conducted merrily, like a children's game. A single characteristic piece of imitation is enough. Some children fail to respond merely on account of timidity or bad humor.

[1] It may be necessary for some children to use a more familiar food, *e.g.*, a piece of cookie, candy, etc.

No. 7. Cognizing real objects by name.

(*a*) Test *S*'s ability to touch the parts of his body that you name. Try easy requests, such as "Where is your hair?" (eyes, nose, ear, etc.) If these are successful, try a few harder questions (such as heart, eye-brow, elbow, etc.). It may be found necessary to give *S* a few illustrative answers to make him understand what is wanted.

(*b*) Have the cup, key and piece of string on a table. Lead *S* to the table, get his attention upon the objects, and say: "Give me the cup," etc. The results may be deemed satisfactory if *S* succeeds in designating the required objects correctly, even if he is somewhat awkward in his manner of doing it. Some defectives will designate an object, but not the right one: others may designate the right object, but will acquiesce if *E* points to another and says: "Isn't this it?"

(*c*) As a first test of suggestibility (see No. 13), ask *S*: "Give me the button." Note carefully his response. Does he pick out one of the three objects; does he hesitate and hunt about the table; does he become confused, or does he promptly say: "There isn't any button"?

No. 8. Cognizing objects in a picture by name.

(*a*) Show *S* the colored picture of The Gardeners. Ask him to point out various objects, *e.g.*, "Where is the ladder? Where is the rake"? Try similarly the picture of the Butcher's Shop, asking "Where is the dog"? "Where is the knife"? etc. Note all errors, but do not correct *S*. Defective children often work zealously at this test; they may, for instance, point out objects before the question is finished, and they seem, in general, unable to adopt a critical or cautious attitude. Imbeciles, it is to be noted, will rarely confess ignorance, will rarely say: "I don't know."

(*b*) For the second part of No. 13, try *S*'s suggestibility by asking: (1) "Where is the *patapoum?*" and (2) "Where is the *nitchevo*"? Note *S*'s attitude and response carefully to see whether he displays any 'resistance' to the suggestion.

No. 9. Naming objects designated in a picture.

Use the colored pictures of the Barber and of the Teamster, and proceed by a method the inverse of that in Test 8, *i.e.*, point to different objects, and say: "What is that"? or ask "What is the man doing"? etc. Use the same questions for all *S*'s and record the answers to each question.

This test is sometimes much more difficult than the 8th, because of the childish vocabulary, faulty pronunciation, or reluctance to talk that some *S*'s display. Tests 7, 8, and 9 represent the same intellectual level—a level which constitutes the approximate boundary between idiocy and imbecility.

No. 10. Comparison of supraliminally different lines.

The purpose here is not to test the fineness of *S*'s discrimination of line-lengths (as in No. 21), but to see whether *S* understands how to make any comparison whatsoever.

(*a*) Show *S* the sheet containing the 30 and the 40 mm. lines, and ask him which line is the longer. Try several times, in chance order, with the longer line now on the right, now on the left, until sure whether *S* can designate the longer line or not. Defectives are apt to answer fairly promptly, but repeated trials may show that they are not really comparing the lines.

(*b*) For the third part of No. 13, now substitute, without *S*'s knowledge, the sheet on which both lines are 40 mm. long, and ask as usual: "Which line is longer now?" Give this "line-trap" test three times.

No. 11. Auditory memory for three digits.

Secure the best attention that *S* can give. Direct him: "Say this after me." Give such series as 3, 0, 8 or 5, 9, 7. Pronounce the digits distinctly, without rhythm, at the rate of 2 per sec. Record *S*'s responses. Many *S*'s fail to understand the instructions; it will then be necessary to illustrate once or twice. The outcome is considered satisfactory if *S* succeeds on the second trial. *E* should note particularly cases in which more than 3

digits are given by S, or in which digits are recited in their natural order, *e.g.*, 3, 4, 5 for 3, 0, 8, since these responses are usually indicative of defective mentality. In practise, this test is combined with No. 19.

No. 12. Comparison of supraliminally different weights.

This test has a similar purpose to that of No. 10. The essential question is: does S know how to go about the task of comparing two weights?

Place before S, about 5 cm. apart, the cubical boxes that weigh 3 and 12 g., and say: "Here are two little boxes: which do you think is heavier?" Try similarly the pairs 6 and 15 g., and 3 and 15 g. Note carefully the procedure adopted by S: does he make a true comparison; does he lift both boxes in the same hand, or does he pass judgment without touching the weights at all? If the procedure is wrong, E should try to explain to him how the weights should be lifted (according to Binet, one in each hand simultaneously). Note whether he succeeds after this instruction.

No. 13. Suggestibility.

The directions for testing suggestibility are given in the three catch-tests already described, viz., the "button trap" (No. 7c), the "word traps" (No. 8b) and the "line trap" (No. 10b). S's that fall into all three 'traps' are extremely suggestible: those that resist all three have no suggestibility[1]: intermediate degrees are indicated by the amount of resistance to the several suggestions.

No. 14. Definitions of familiar objects.

The purpose here is to ascertain S's ability to put a very simple idea into words, by asking him to define the words:—*house, fork, mamma.* First make sure that S knows the word, then ask him to define it; thus, "You know what a fork is, don't you?" "Yes." "Well, what is it?" If S 'balks' at the definitions, illustrate by a

[1] So far as these tests go: for other methods of measuring suggestibility, consult Chapter X., Tests 40-44.

simple definition, *e.g.*, of a dog as "an animal that barks." Record *S*'s replies verbatim. Try all three words. Watch for absurd replies, especially for "echoing" replies, as "A house—why it's a house." Most definitions will be given in terms of function, *e.g.*, "A fork is a thing to eat with."

No. 15. Memory for sentences.

This test resembles No. 11, save that sentences of approximately 15 words are substituted for the three digits. Use the following eight sentences in the order given.[1] Speak slowly and distinctly. Do not repeat any sentence if *S* fails at the first trial. As in other tests, note particularly the presence of absurdities in *S*'s responses: a mere omission is but an error of memory; an absurdity is an error of judgment and is probably symptomatic of mental deficiency.

(1) I get up in the morning, play all day, and go to bed at night.
(2) The horse is drawing the wagon; the road is steep and the wagon is heavy.
(3) In summer the days are fine; in winter it snows and I am cold.
(4) Mabel has been naughty; she will get a scolding because she didn't mind her mother.
(5) It is almost one o'clock; the house is quiet and the cat is sound asleep.
(6) The apple tree casts a pleasant shade on the ground where the children are digging.
(7) It is not necessary that you should tell everything that you happen to hear about.
(8) We should be careful not to confound the critical spirit with the spirit of contradiction.

No. 16. Differences between familiar objects recalled in memory.

This test is designed to study *S*'s capacity to observe, to recall clearly, and to draw distinctions. Three pairs of objects are contrasted—(*a*) paper and cloth, (*b*) a fly and a butterfly, (*c*) wood

[1] These sentences differ somewhat, both in content and in order, from the originals, on account of the impossibility of preserving in translation the characteristics of the French sentences. Slightly variant renderings will be found in Goddard. See No. 2 of the 1908 series (Test 54) for another set of test sentences.

and glass. In each case, make sure first that *S* is familiar with both objects, then ask him to state the difference between them. For example: "Do you know what paper is? . . . Do you know what cloth is? . . . Then tell me the difference between them." It is usually necessary to indulge in further questioning, *e. g.*, "Why aren't they the same?" "How can you tell them apart?" etc. If any of these pairs is unknown, *E* may try other objects, such as a plate and a cup, milk and water, etc. In general, three types of responses appear: (1) some *S*'s cannot be brought to understand the idea of expressing differences; (2) some *S*'s make absurd answers, *e.g.*, "A fly is larger than a butterfly"; (3) other *S*'s succeed in making a satisfactory distinction.

No. 17. Memory for pictures.

Instruct *S*: "I'm going to show you some pictures, and let you look at them for half a minute. After that I shall ask you to tell me the names of all of the pictures you saw." Display the first card (mouse, bag, bucket, etc.) for 30 sec. If *S* appears inattentive, remind him of his task: "Don't forget, look at them all," etc. Record the number of objects named; note the omissions, insertions, and other mistakes. Note also *S*'s general manner of procedure—whether his attention seems keen and persistent, whether he pronounces the names aloud, etc. If the first trial fails for any reason, try the second card of objects. If time permits, try both groups with all *S*'s.

No. 18. Drawing from memory.

Instruct *S*: "I'm going to show you two drawings (Fig. 59) now, and let you look at them for 10 sec. After that I want you to draw them from memory as well as you can." Most *S*'s need a little encouragement in this test. Care should be taken that they begin work promptly. The results may be scored, roughly, as exact, similar, or very unlike (Binet and Simon), or they may be graded on a scale of 10 for satisfactory reproduction (Decroly and Degand).[1]

[1] For reproductions of the drawings made in this test by defective children, consult Decroly and Degand, p. 111.

No. 19. Auditory memory for more than three digits.[1]

Follow the directions for No. 11, but use test-numbers like the following: 5328, 41053, 708219, 2690325, 85023908. In each case, the digits are to be pronounced successively, as: five, three, two, eight, etc. Give three trials for each kind of number: 4-place, 5-place, etc., until a number is reached of such a length that *S* is successful in no one of three trials. Let *S* state after each repetition whether he thinks it is correct or not.

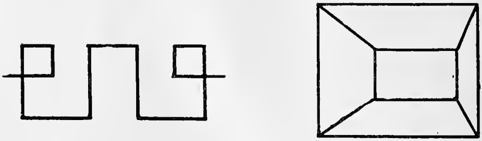


FIG. 59. DESIGNS FOR BINET-SIMON TEST NO. 18, 1905 SERIES.

No. 20. Resemblances between familiar objects recalled in memory.

This test is similar to No. 16, but here *S* is asked for points of likeness instead of points of difference. The same general precautions are to be taken. It will often be necessary to explain that two different objects may have some points of resemblance. The three groups of objects employed by Goddard are: (*a*) milk and snow, (*b*) a mosquito and a bee, (*c*) a table, a chair and a door.

No. 21. Discrimination of lines.

The procedure resembles that of No. 10, but the line-lengths are now only liminally different.

Show *S* the test-cards with the longer line now on the right, now on the left, until he has had four trials with each card. Use first the cards with the 30 mm. standard, then those with the 100 mm. standard. Let *S* judge in each trial whether the right-hand line is longer or shorter. The discrimination is difficult even for a normal adult *S*.

[1] Compare Test 38 and the results there given for normal children.

No. 22. Arrangement of five weights.

Place on the table the cubical boxes weighing 3, 6, 9, 12, and 15 g. Say to *S:* "Here are five little boxes; they look alike, but they don't weigh the same. I want you to arrange them in order. First put here [at the left] the lightest of all, next the one that is just a little heavier, then the next heavier, then the next heavier, and here [at the right] the heaviest of all." This explanation must be tried in several forms, as it is rather difficult to make intelligible to many *S*'s, but the explanation must not suggest that the weights are to be lifted. Note *S*'s manner of procedure: does he understand the directions, does he test the weights—with one hand or both hands, etc.? Record the order of the weights and rank *S* in terms of the number of displacements, *i.e.*, the number of changes requisite to bring the weights into their proper order. Mix up the weights and let *S* try a second, and a third time. Compute the total number of errors for the three trials. The computation of errors may be illustrated as follows: if the order is 3, 6, 9, 15, 12, there are two errors; if the order is 9, 15, 6, 3, 12, there are eight errors; if 3, 9, 6, 15, 12, there are four errors.

No. 23. Detection of the missing weight.

If *S* succeeds fairly well in the preceding test, let him close his eyes while *E* picks out the 9-gram weight. Let *S* then examine the four remaining weights, which have, meanwhile, been placed equidistantly, to see whether he can, by lifting alone, tell which weight has been removed. Repeat the test by the removal of the 6-gram weight, and again by the removal of the 12-gram weight, leaving four weights on the table in each instance.[1]

No. 24. Rimes.

Making rimes is employed to test extent of vocabulary, general 'nimbleness' of mind, spontaneity, etc. *E* must first discover whether *S* knows what a rime is. If he does not, *E* should try to

[1] This procedure is preferable, for comparative purposes, to the removal of a weight by chance as practised by Binet and Simon.

explain by simple illustrations, as *dog*, *fog*, *log*, etc. If the explanation succeeds, try for a simple test the naming of rimes to the word *ball*—allowing one minute for the trial. Most *S*'s that understand the test will be able to name such common words as *fall*, *call*, *all*, *small*, etc. Next try the word *coy*, then the words *din*, *feet*, *spring*, *money*.[1] Record in each case the words suggested by *S* in one minute.

No. 25. Missing words.

The method resembles that of the Ebbinghaus 'completion' method (see Test 48), but is simpler in that only the final word in each sentence or phrase is omitted and the test is conducted orally. In general, the difficulty of such a test depends almost entirely upon the nature of the 'gap' that is to be filled. For that reason, all *S*'s should be given the same test-sentences, spoken in the same manner, viz: with the voice suspended at the missing word (the words in parentheses are omitted). If necessary, *E* must urge *S* by saying: "Finish it," "What comes next?" or "What should I say next?" etc.

"One day a dog was walking along with a large piece of meat in his (mouth). Pretty soon he came to a brook and began to cross it on a (bridge). As he looked into the water, he thought he saw another dog with another piece of (meat). I'll have that piece, too, said he, dropping his (own). But his meat sank out of (sight). He saw nothing but his own shadow in the (water). Then he knew that he had been too (greedy). I hope next time he will be a wiser (dog)."[2]

No. 26. Sentence-building.[3]

Ask *S* to make a sentence containing the three words *girl*, *river*, *ball*.[4] Since many *S*'s do not understand the meaning of the term sentence, *E* must give illustrations with other words, and try various forms of explanation, *e.g.*, "Tell me something and use all these

[1] These are the test-words employed by Goddard.

[2] This is substituted for the selection used by the French investigators. Other words than those indicated may, of course, be properly substituted, *e.g.*, *plank* for *bridge*.

[3] For similar tests, consult Test 46, and No. 43 of Test 54.

[4] These terms are substituted for those of Binet and Simon.

words when you tell it to me,"etc. Many S's fail entirely; others are able to construct sentences containing one or two of the words; others make three sentences with one of the words in each. Record the results verbatim.

No. 27. Replies to problem-questions.

This important test is introduced to discover S's ability to pass a satisfactory judgment as to what should be done in certain situations: it is, in short, a test of the 'problem-solving' type. The questions are to be propounded orally in the order given below:[1] the replies are to be recorded verbatim, and graded in terms of the appropriateness and degree of 'common-sense' that they display.[2] E must exercise patience, give S plenty of time, and avoid suggesting failure by such remarks as: "You don't know, do you?" The slowness born of caution or difficulty in formulation of the reply must not be confused with that born of ignorance.

(1) What's the thing to do when you feel sleepy?
(2) What's the thing to do when you feel cold?
(3) What's the thing to do when you think you'll be late to school?
(4) If you find it's raining, when you start to leave your home, what's the thing to do?
(5) If you are out walking, and feel tired and haven't any money to pay for a ride, what's the thing to do?
(6) What's the thing to do when you miss a train?
(7) What's the thing to do when you break something that doesn't belong to you?
(8) What's the thing to do if you find that somebody has stolen a book out of your desk at school?
(9) What's the thing to do if you find out that your house is on fire?
(10) What's the thing to do if the boy you are playing with hits you without meaning to?
(11) What's the thing to do when you want good advice?
(12) What happens to a person who is lazy and doesn't want to work?

[1] In practise, the French investigators seem frequently to have used only the first 20 questions.
[2] Some experience is required, in grading the answers, to avoid either underestimation or overestimation of their real value.

(13) Why should a person save some of his money instead of spending it all?

(14) What would you do if you were punished when you didn't deserve it?

(15) Suppose a person is planning to do something important, what's the first thing for him to do?

(16) Suppose you wanted ten cents, how would you earn it?

(17) Suppose somebody has done something you didn't like and comes to say he is sorry, what's the thing for you to do?

(18) Suppose somebody asks you what you think about a boy that you don't know very well, what's the thing for you to do?

(19) What happens when two persons discuss a question without understanding the words?

(20) Suppose a boy always contradicts you, whatever you say, what's the thing for you to do?

(21) Why should you judge a person by what he does rather than by what he says?

(22) Why should you forgive an injury done in anger more quickly than you would forgive one done without anger?

(23) Why is it better to go on and persevere with what you have begun than to begin something else?

(24) Why shouldn't you keep reminding a person of the things you have done for him?

(25) If you have committed some injury that can't be made right again, what's the thing to do?

As illustrations of the grading of answers, take Question 9; a good answer would be : "Throw water on it if you can, then send in a fire alarm." Less good: "Save yourself and whatever else you can." Absurd answer: "Put stewpans on the fire and open the windows to let the smoke out."[1]

No. 28. Interchange of the clock-hands.

Here it is proposed to demand of *S* active attention, visual imagination, and simple reasoning. Let *E* first make sure that *S* knows how to tell time. Imbeciles are prone to declare that they can, but to fail utterly on trial. If *S* can tell time, propound the following query: "Suppose it is four minutes of three, can you see in your mind where the large hand would be, and where the small

[1] For fully illustrated accounts of answers to all questions, consult Binet and Simon, pp. 278–292, and Decroly and Degand, pp. 120–1.

hand would be?" If the answer be affirmative, continue: "Now suppose the large hand takes the place where you say the small hand is, and the small hand takes the place of the large hand, then what time would it be?" Try similarly the interchange of the hands when at twenty minutes past six. *E* should be satisfied to receive approximate answers, viz: 11.15 and 4.30, but if these are given successfully, *E* may say: "These answers are not absolutely right, though very nearly so; can you tell me why they are not exactly right?"

The test is difficult for most *S*'s; some cannot undertake it at all or give answers that are utterly wrong; others give answers that are partly right, *e.g.*, one hand is right, or the right and left sides of the dial are confused; still others give the correct answers, but only the very capable can explain the slight discrepancy.

In conducting the test, *S* must not be allowed to look at any clock or watch face, or to aid himself by drawing, but must work the problem out mentally.

No. 29. Drawing from a design cut in a quarto-folded paper.

The solution of this test exacts capacities like those demanded in the preceding test.

Show *S* the two sheets of paper and call his attention to the fact that they are just alike. Leave one sheet on the table, and, while *S* watches carefully, fold the other slowly twice, *i.e.*, first into halves, then into quarters. Then with the scissors cut two triangular sections out of the folded edge of the sheet (small Fig. 1 in Fig. 60), so that, were the sheet unfolded, it would have the appearance indicated (small Fig. 2). Give *S* no opportunity to examine the sections that have been cut out. Put the folded sheet before him, saying: "You see I have cut a piece out of this paper. If I should open it, it wouldn't look like that piece [pointing to the uncut open sheet]. I want you to draw on that piece what we should see if we unfolded this one." Give these directions carefully, no more, no less. If this test proves too difficult, it may be tried in the simpler form used by Binet, in which *E* cuts out only the upper notch of Fig. 1.

In the lower part of Fig. 60 are shown the results obtained by

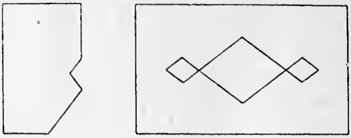


Fig. 1. Fig 2.

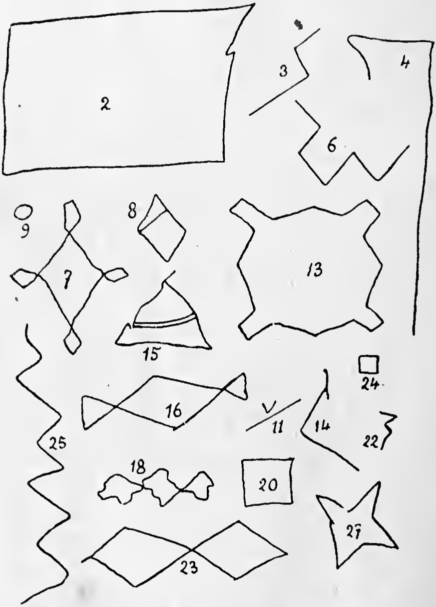


FIG. 60. ORIGINALS AND REPRODUCTIONS OF BINET-SIMON TEST NO. 29.

(From Decroly and Degand.)

Decroly and Degand from different S's. The numbers refer to the numbers of the different S's, but do not correspond with the numbering indicated for the same S's in Table 99, below.

No. 30. Distinction between abstract terms.

Without preamble, ask S: "What is the difference between liking a person and respecting him"? And again, "What difference is there between being vexed and being bored"?

Results and Conclusions.—(1) In general, S's fall into *four groups* with regard to each test; first, those who fail completely; secondly, those who succeed partially; thirdly, those who succeed completely; fourthly, those whose mistakes are of the absurd or ridiculous kind: these last are significant for the diagnosis of mental deficiency, and usually indicate something else than simple ignorance or simple inability due to immaturity.

(2) The chief *obstacles* encountered in the examination of defectives are ignorance of what is wanted, resistance springing from a morose or peevish disposition, timidity, and inability to resist distraction and give continuous attention to the work.

(3) Binet and Simon have established the following *standards for normal children* for the ages 3 to 12.

(*a*) Nos. 1–7 are passed by all children 3 years old and upward.

(*b*) Nos. 8–12 and No. 14 are passed successfully by children 7 years old and upward: 5-year old children need some help in Nos. 9 and 12: 3-year old children make some errors in Nos. 8 and 9, and succeed but rarely in No. 11.[1] The characteristic mark of the 3-year old is therefore the ability to name familiar objects in pictures, of the 5-year old to repeat three numbers, to compare two lines or two weights (after instruction), and to define a well-known object.

(*c*) Suggestibility (No. 13) is extreme in the 3-year child, and decreases progressively in later years: the "button-trap" is avoided by most 5-year olds.

(*d*) The *average 7-year old child* repeats 3 of the 8 sentences and makes 3 mistakes ("absurdities"), recalls 4.3 of the 13 pictures

[1] In the author's opinion, this conclusion is dubious, as many 3-year old children of his acquaintance can repeat three digits correctly.

and can repeat 5.3 digits. In No. 21, he makes 1.5 errors. In No. 22, he has a misplacement of the weights of 11.3. He makes an indefinitely large number of mistakes in No. 23. He cannot make rimes, and fails to pass Nos. 26, 28, 29 and 30. In No. 27, (first 20 questions) he is silent from 6 to 11 times, and makes one or two absurd replies.

(*e*) The *average 9-year old child* repeats 4 of the 8 sentences, recalls 6.2 of the 13 pictures and can repeat 6 digits. In the discrimination of 30 mm. lines he makes 1 mistake. He has an average misplacement of 4 in No. 22, and makes from 2 to 5 errors in the missing-weight test (No. 23). He can make from 1 to 4 rimes, but apparently does no better than the 7-year old child in Nos. 26–30.

(*f*) The *average 11-year old child* repeats 5 sentences with 0.5 error, recalls 7.2 pictures, and can repeat 6 digits. In discriminating 30 mm. lines, he makes 0.2 errors. His misplacement of weights is 2.4, and his errors in the missing-weight test are 2. He may succeed very crudely in drawing from the cutting. He makes several rimes, does fairly well with No. 27 (minimum, 2 "silences" and 0.5 "absurdity," maximum 5 "silences" and 2 "absurdities"), but does scarcely anything with Nos. 26, 28, 29 and 30.

(4) Systematic application of these tests to defective children by Decroly and Degand has yielded the results which are summarized in Table 99.[1] In the light of the explanation which follows the Table, it will be seen:

(*a*) that there is a very good general correspondence between the classification of these defectives on the basis of pedagogical, clinical, and general observation and their classification on the basis of the tests, so that the tests have a practical value for the psychological classification of defectives;

(*b*) that they are not well adapted for the diagnosis of deaf-mutes (or, it may be added, of moral defectives);[2]

[1] Experimenters who wish to apply these tests to defectives are recommended to consult the detailed results from which this summary has been made, as well as the results of Binet and Simon with abnormal children.

[2] Decroly and Degand suggest as improvements for the Binet-Simon series, the addition of tests adapted for deaf-mutes and moral defectives, the addition of tests bearing more particularly upon what they term "active intelligence, alertness, or logic in action" to fill out the evident inadequacy in this respect of Nos. 18 and 22, and the rearrangement of certain tests that at present are unnecessarily long, or that entail needless repetition.

TABLE 99

Application of the 1905 Binet-Simon Tests to Defectives (Decroly and Degand)

No. of child	Number of Test 6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	Y	Y																							
2	Y	Y	Y	S																					
3	Y	Y	Y	Y	P			P	S	S		S													
4	Y	Y	Y	P	P	Y	P	P	P	P	P	P	P	P	P	P									
5	Y	Y	Y	Y	Y	Y	Y	S		S		S	S			S						S			
6	Y	Y	Y	S	Y	Y	Y	P	S	S	Y	S	S	S	S	S	P	P	S	S		S			
7	Y	Y	Y	S	Y	Y	S	P	S	S		S		S											
8	S	Y	S	S	Y		Y						S			S								S	
9	Y	S	S	S	S	Y	Y	S	S	S	S	Y	Y	S	S	S	Y	Y		S		S			
10	Y	Y	Y	Y	Y	Y	Y	S	S	S	Y	Y	S	S		S	S	S				S	S		
11	Y	Y	Y	Y	Y	Y	Y	P	S	S	S	Y	S	S	S	S	Y	Y	S	S		S			
12	S				Y								Y											S	
13	Y	Y	Y	S	Y	Y	Y	S	Y	S	S	Y	S	S	S	P	Y	Y	Y	Y	P	S	S	S	P
14	Y	Y	Y	S	Y	Y	Y	S	Y	Y	Y	Y	S	Y	S	Y	Y	Y	Y	Y	S	Y	P	S	P
15	Y	Y	Y	S	Y	Y	Y	P	S	Y	S	S	S	S	S	Y						S		S	
16	Y	Y	Y	S	Y	Y	S	Y	Y	S	Y	S	S	Y	S	S	S	S	Y	S	P	S	S	S	P
17	Y	Y	Y	Y	Y	Y	S	S	Y	Y	Y	Y	Y	S	Y	S	S	S	Y	Y	P	S	Y	S	P
18	S	Y			Y							S				S								S	
19	S	Y	Y		Y		Y	S					S			Y	Y	Y						S	
20	Y	Y	Y	Y	Y	Y	Y	S	Y	S	Y	Y	Y	S	S	S	Y	Y	Y	Y		S		Y	P
21	S	Y	Y	S	Y		S	S				S				Y	Y	Y						Y	
22	Y	Y	Y	Y	Y	Y	Y	S	Y	S	Y	Y	S	S	S	Y	Y	Y	S	S	S	S		Y	
23	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	S	Y		S	S
24	Y	Y	Y	Y	Y	Y	Y	Y	S	Y	Y	S	Y	Y	S	S	S	Y	Y	S		S	Y	Y	P
25	Y	Y	Y	Y	Y	Y	Y	S	Y	Y	Y	Y	Y	Y	S	Y	S	Y	Y	Y	Y	S	S	S	S

Explanation of the table. The symbol Y (Yes) indicates complete success; the symbol S (Some) indicates partial success; the symbol P (Poor) indicates poor results; omission of a symbol indicates no response at all. It must be understood that many of the omissions in this table are cases in which the tests could not be applied; thus, Nos. 8, 12, and 21 are deaf children, while Nos. 18 and 19 are partially deaf. The first 5 tests are passed by all.

The children are numbered in order of their general intellectual capacity: Nos. 1 and 2 exhibit extreme defect, and are classed as idiots with imitation of gestures; Nos. 3 to 7 exhibit marked defect, and are classed as imbeciles; Nos. 8 to 13 are feeble-minded children with moderate defect; Nos. 14 to 17 suffer only mild deficiency and may be ranked as backward children; Nos. 18 to 25 possess normal intellectual capacity in general, but need, with the exception of No. 25, a certain amount of assistance in order to succeed outside of an institution.

(*c*) that the tests are not so serviceable for the diagnosis of children whose capacity lies between the normal and the abnormal;

(*d*) that the tests are not arranged at present in precisely the order of their difficulty, at least for defective children, *e.g.*, No. 9 is evidently more difficult than No. 10; or, in other words, that failure to pass a given test does not necessarily involve failure to pass all subsequent tests, *e.g.*, Child No. 13.

(5) Binet and Simon advance the following classification of defectives on the basis of their tests: the figures in parentheses indicate the number of the limiting test.

A. Idiots (no use of language).

A 1. Vegetative idiot with no relational activity (0).
A 2. Idiot with visual coordination (1).
A 3. Idiot with prehension (2, 3).
A 4. Idiot with knowledge of food (4).
A 5. Idiot with capacity to imitate (6).

B. Imbeciles (roughly equal to the 2–5 year normal child).

B 1. Imbecile with capacity to name (7, 8, 9).
B 2. Imbecile with capacity to compare (10, 11, 12).
B 3. Imbecile with capacity to repeat sentences (15).

C. Feeble-minded.

C 1. Feeble-minded with capacity to state differences (16).
C 2. Feeble-minded with capacity to make serial arrangements (22).

REFERENCES

(1) A. Binet and T. Simon, (*a*) Sur la nécessité d'établir un diagnostic scientifique des états inférieurs de l'intelligence, in A. P., 11:1905, 163–190. (*b*) Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux, *ibid.*, 191–244. (*c*) Application des méthodes nouvelles au diagnostic du niveau intellectuel chez des enfants normaux et anormaux d'hospice et d'école primaire, *ibid.*, 245–336.

(2) O. Decroly and (Mlle) J. Degand, Les tests de Binet et Simon pour la mesure de l'intelligence: contribution critique, in Ar. P., 6: 1906, 27–130.

(3) H. H. Goddard, The Binet and Simon tests of intellectual capacity, in The Training School, 5: December, 1908, 3–9. (Also reprinted as "The grading of backward children,' New Jersey Training School, Dept. of psychological research, August, 1909.)

TEST 54

Binet-Simon graded tests: 1908 series.—The following tests, like the preceding group by the same authors, form a graded series for the quantitative determination of native ability and general intelligence. Many of them are repetitions, with slight modifications, of the preceding tests, but there are 33 new tests, so that the series as a whole is more extensive, and may be assumed to represent a more satisfactory combination for the purpose, not only of diagnosing the degree of retardation of backward children, but also of estimating the degree of development of normal children from 3 to 13 years of age.[1]

The tests have been classified by the ages to which they apply in normal children. An effort has been made to eliminate the variable error due to training, by giving preference to tests which measure native ability or natural acquisition rather than erudition or the capacities developed by school training.

As before, the tests are to be administered individually, with suitable precaution to avoid timidity or obstinacy on the part of the subjects. *E* should encourage *S* by the adoption of a friendly and kindly tone of voice, but should not give him direct assistance or instruction in explanation of the tests (save where noted), and should never venture to criticize. He must be tactful, patient, and sympathetic. It is well to have a third person present to act as clerk for recording *S*'s answers, but no spectators should be allowed.

[1] As it is not clear whether Binet and Simon regard this as a complete substitute for their earlier series, the tests have been reproduced entire, following the divisions of the authors into groups which are intended to represent norms of average capacity for the several ages mentioned. These performances are those of more than 200 French school children, drawn in large degree from the poorer classes. It is to be desired that norms may soon be established for American children with the same series of tests, especially since the recent investigation by Decroly and Degand of normal children in the Brussels schools makes it evident that not a few of the Binet-Simon age-norms must be revised.

Experimenters will doubtless find it profitable to select some tests from the 1908, and some from the 1905 series, because, as Decroly and Degand point out, several of the tests now omitted were of special interest.

A. TESTS FOR 3-YEAR OLD CHILDREN

No. 1. Comprehension of language.

This is identical with No. 7 a of the 1st series.

No. 2. Memory for sentences.

This test is like No. 15 of the 1st series, save that the sentences are arranged in a progressive series of 2, 4, ... 42 syllables.[1]

The sentences are to be tested in the order given (though the shorter ones may be omitted for older children). Each sentence is read aloud to *S* slowly and distinctly, and he is asked to repeat what he has just heard.

The average 3-year old can repeat a combination of 6 syllables (No. 3), but not one of 10 syllables (No. 5).

(1) Papa.
(2) Slipper. Letter.
(3) It is cold and snowing.
(4) I have a dog. He's a fine one.
(5) His name is Jack. Oh, what a naughty boy!
(6) It is raining outdoors, but we can stay inside.
(7) We are having a fine time. We found a mouse in the trap.
(8) Let's all go for a walk to-day. Please give me that big hat to wear.
(9) Poor Helen has just torn her new dress. She will surely feel sorry for that.
(10) Why should any one want to do injury to such beautiful creatures as birds?
(11) We expect to have a great time at the seashore, digging in the white beach sand all day long.
(12) When the train crosses the road the engineer will blow the whistle and the fireman will ring the bell.
(13) My young brother Frank had a fine time on his vacation this summer: he went fishing almost every day.
(14) To start a fire in the open is one of those tricks that everyone thinks he can perform until he tries it.
(15) He sinks the net in the water and waits until he can see the fish distinctly, lying perfectly still and within reach.

[1] The shorter sentences are largely adaptations of the series used by Binet and Simon, and the general type and subject-matter of the original sentences has been preserved so far as possible. The longer sentences have been arranged by the author. This test is not, of course, intended to be used in its entirety for 3-year old children: the longer sentences are used for older children, as is described below. The whole series has been collected in one group for convenience in reference.

(16) The first rapid was only the beginning: half a mile below we could see the river disappear between two points of rock.

(17) One day the children's grandmother came to visit them and the children were very happy because she told them beautiful fairy stories.

(18) Louis looked out of his window one night and saw that some of the older boys were burning big piles of dry autumn leaves in the street before his house.

(19) The Indians used to perform many kinds of queer, horrible dances in the course of which they yelled and shrieked as if suffering the most painful torture.

(20) A full-grown grizzly bear will usually weigh from five to seven hundred pounds, but exceptional individuals undoubtedly reach more than twelve hundred-weight.

(21) When you set out to explore one of these minor streams in your canoe, you have no intention of epoch-making discoveries or thrilling or world-famous adventures.

No. 3. Memory for digits.

This is the same as Nos. 11 and 19 of the 1st series, save that *S* is tested with one, and with two, as well as with three or more digits. Many 3-year old children can repeat 2 digits without difficulty, but cannot repeat 3 digits correctly.

No. 4. Description of a picture.

This test resembles somewhat both the description test of Chapter VIII, and No. 9 test of the 1905 series, though it differs from both in some essential features. The authors regard it as extremely valuable, because it arouses great interest, and because it affords insight into the grade of *S*'s intelligence.

MATERIAL.—Set of 8 colored pictures, as in Nos. 8 and 9 of the 1905 series.[1]

[1] Dr. Goddard, who has employed these pictures in the examination of retarded children at Vineland, N. J., writes of this test: "We find it valuable to use the whole set of pictures, or at least five or six of them, because it often happens that a child who does not at first notice the action [*i.e.*, does not appreciate or name the situation that is pictured] because his attention is attracted to some special feature, may notice it in the 2d, or 3d, or 4th, or even sometimes not until the last picture. If we get this reaction at all, we feel that the child displays a higher grade of intelligence than if he merely names a man or a woman or some other single object. We find many intermediate grades that are very suggestive."

METHOD.—The picture is shown to *S*, who is urged to describe it (from direct observation). *E* may ask: "What is that"? "What do you see there"? "Tell me what that is about," etc.

RESULTS.—At least three types of answers may be distinguished: (1) simple enumeration, *e.g.*, "A man, a dog, a knife," etc., (2) description, (3) interpretation. The 3-year old child gives only replies of the first type, though the best children at this age may use a few connectives, *e.g.*, "A man and a dog," etc.

No. 5. Family name.

Ask *S*: "What is your name"? If only the Christian name is given, try to ascertain whether the child also knows his family name. Many 3-year old children fail to pass this test.

B. TESTS FOR 4-YEAR OLD CHILDREN

No. 6. Own sex.

Ask *S*: "Are you a little boy or a little girl"? Many 3-year old children answer erroneously, but the normal 4-year old child does not.

No. 7. Naming familiar objects.

Show *S* successively (1) a key, (2) a closed penknife, (3) a penny: in each case ask: "What is that"? According to Binet and Simon, most 3-year old *S*'s, but only a few 4-year old *S*'s, fail.

No. 8. Memory for three digits.

This is the same as No. 3. Four-year children can repeat 3 digits.

No. 9. Comparison of two lines.

This test is closely similar to No. 10 of the 1st series, save that the authors employ slightly longer lines, viz: a 5 cm. and a 6 cm. line, drawn parallel to one another at 3 cm. distance.

Normal 4-year old *S*'s should be able to indicate the longer line without hesitation.

C. TESTS FOR 5-YEAR OLD CHILDREN

No. 10. Comparison of supraliminally different weights.

This test is identical with No. 12 of the 1st series. The actual weighing is, of course, not the difficult feature: the question is: does *S* know how to make a comparison of two weights? The test can, as a rule, be passed at 5 years, but not at 4.

No. 11. Copying a square.

Draw with ink a square about 4 cm. on a side, and ask *S* to reproduce this square. He must use pen and ink (or fountain-pen), but not a pencil. It will be found that most children reduce

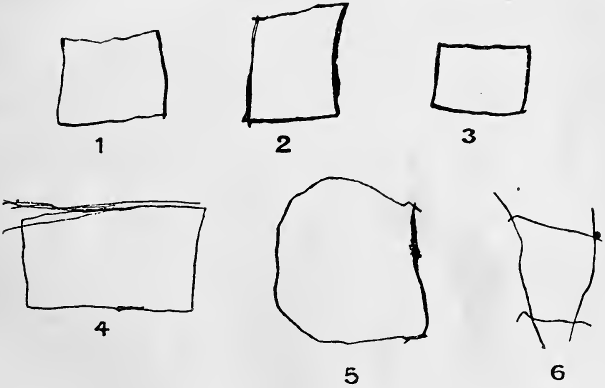


FIG. 61. COPIES OF THE SQUARE.
(No. 11, 1908 Binet-Simon series.)

the size of the square, but their drawing is to be graded entirely in terms of its qualitative excellence. The first three drawings in Fig. 61 are ranked as successful, the last three as unsuccessful attempts at reproduction.

No. 12. The divided rectangle.

Provide two bits of cardboard, each 8 × 14 cm. Cut one card along its diagonal. Place upon a table the uncut card and the two

halves of the cut card, so disposed that their hypotenuses are not in juxtaposition. Ask *S:* "Put these two pieces together to make a figure like that one" [pointing to the rectangle]. If *S* balks at the test, he *must* be encouraged to try: if he happens to turn one piece over, he must be started again: if he asks questions as to the correctness of his solution, *E* should not venture assistance.

Not more than one in 3 of 4-year old *S*'s succeed, but not more than one in 12 of 5-year old *S*'s fail.

No. 13. Counting four pennies.

Place on the table four pennies, close together, all in full view. Say to *S:* "How many pennies are there? Count them." Insist upon distinct counting—"one, two, three, four."

No 3-year old *S*'s succeed; about half of the 4-year olds succeed, but at 5 years, only backward children fail.

D. TESTS FOR 6-YEAR OLD CHILDREN

No. 14. Knowing right and left.

Ask *S:* "Show me your right hand," and, after this is done, "Show me your left ear." Be careful to give no hint by word or look. The test is passed even if the responses are made hesitatingly. Practically every 4-year old child will point to his right instead of to his left ear: about one child in three of the 5-year olds makes an error, whereas no child fails at the age of 6.

No. 15. Memory for sentences (16 syllables).

Children of six years should be able to repeat the first 9 sentences of Test No. 2. Only one-half of those tested at the age of 5 can accomplish this.

No. 16. Elementary esthetic judgment.

S is shown in succession a card containing the three pairs of drawings of women's heads shown in Fig. 62, and is asked in each instance: "Which of these two is the prettier"? The correct answers are given by 6-year old children, but by only about one half of 5-year old children.

FIG 62. DRAWINGS FOR TESTING ELEMENTARY ESTHETIC JUDGMENT.

(No. 16, 1908 Binet-Simon series.)

No. 17. Definitions of familiar objects.

The test is the same as No. 14 of the 1905 series, save that the words to be defined are (1) *fork,* (2) *table,* (3) *chair,* (4) *horse,* (5) *mamma.* Practically all children six years old succeed in giving simple functional definitions, *i.e.,* definitions in terms of use; younger children usually define only one or two of the terms.

No. 18. Execution of a triple order.

Young children can easily execute a single simple order (see No. 6 of the 1905 series), but may fail if the order is complicated. Give *S* an order which really combines three separate orders in the one set of instructions, for example: "Here is a key; please put it on that chair. Then please close the door. Then you'll notice a box on a chair near the door; please bring me that box. Do you understand? First put the key on the chair, then close the door, then bring the box. Now, go ahead."

Hardly any 4-year old child executes all three orders correctly; of 5-year old children, about one-half succeed; of 6-year old children, nearly all succeed.

No. 19. Own age.

Ask *S:* "How old are you"? It is only from the 6th year up that the majority of *S*'s answer correctly: prior to that time, most children give their age too young.

No. 20. Knowing morning and afternoon.

Ask *S:* "Is it morning or afternoon"? Since a tendency is shown by some children always to take the latter of two alternatives, it is well, if the time is afternoon, to put the question: "Is this afternoon or morning"? Young children, if they answer at all, answer

more or less evidently by chance, but most 6-year children answer correctly and without hesitation.[1]

E. TESTS FOR 7-YEAR OLD CHILDREN

No. 21. Unfinished pictures.

S is shown a card containing the four outline drawings of Fig. 63, and is asked, with respect to each drawing: "What is gone (lacking) in that picture"? If he replies "legs," "neck," etc., though these answers are, strictly speaking, correct, he is urged to say whether there is anything else missing. The test is passed if the correct answer is given in 3 of 4 cases.

The replies of 5-year old *S*'s are usually entirely inadequate; those of 6-year *S*'s are correct in about one-third of the cases, but those of 7-year *S*'s should be correct in three-fourths of the cases.

No. 22. Number of fingers.

Ask *S*: "How many fingers have you on your right hand"? "On your left hand"? "How many does that make on both hands"? Pause for the reply after each question. The replies must be made quickly, without stopping to count, and all three questions must be answered correctly to pass the test.

This is essentially a 7-year test, since half those tested at 6 years fail.

No. 23. Writing from copy.

The phrase "See little Paul" is written in ink. *S* is to copy this with pen and ink. The test is passed if the copy is sufficiently legible to be read by a person who did not know what was to be written. This test is, of course, conditioned by training, but nevertheless it forms a useful index of general intelligence.

[1] It may be well to point out that in this, and in many of the tests, the grading of average ability seems at first sight too low, but due allowance must be made for the fact that the child is placed under rather unusual conditions and that the tests are made rapidly, so that confusion may easily appear.

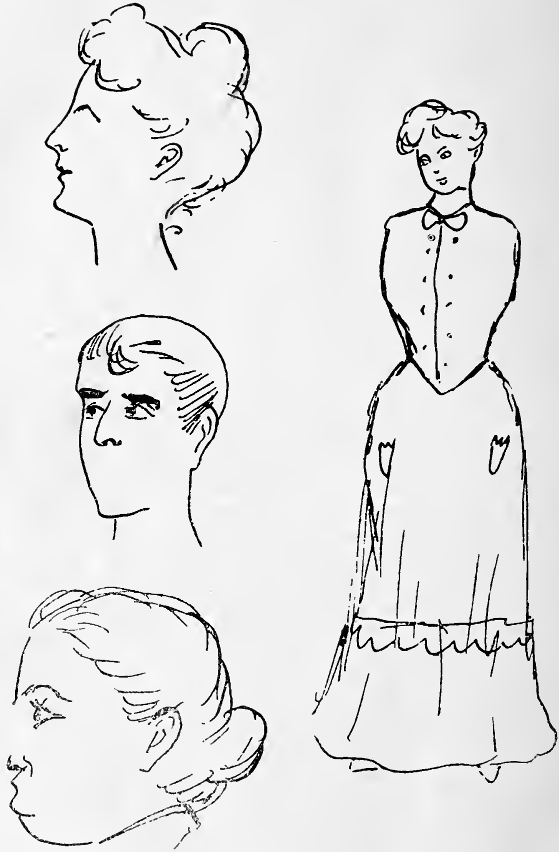


FIG 63. THE UNFINISHED PICTURES.
(No. 21, 1908 Binet-Simon series.)

No. 24. Copying a diamond.

Repeat No. 11, but use a diamond instead of a square.

No. 25. Memory for five digits.

This is the same as No. 3. Seven-year *S*'s should repeat 5 digits.

No. 26. Description of a picture.

This is the same as No. 4. Seven-year children should not merely enumerate individual features, but describe the picture as a scene.

No. 27. Counting thirteen pennies.

The test is identical with No. 13, save that 13 pennies are substituted for the 4 there employed. The pennies, which are placed in a row, must be counted with the finger, aloud, without an error.

No. 28. Naming four common coins.

S is shown, successively, a penny, nickel, dime, and quarter, and asked to name each one.[1] Passed by a majority at the age of 7.

F. TESTS FOR 8-YEAR OLD CHILDREN

No. 29. Reading and report.[2]

Binet and Simon, as noted in the previous series, characterize the imbecile by inability to use printed or written language (after an effort has been made to teach him, of course). The following test is, therefore, deemed by them important as a boundary test.

METHOD.—Ask *S* to read aloud the following, and record his time in sec.

[1] Binet and Simon used 0.05, 0.10, 0.50, and 5.0 franc pieces.

[2] See Chapter IX, Test 39, for more elaborate tests of 'logical' memory, as applied to normal children, and Chapter VII, Test 28, for other tests of reading.

Three Houses Burned.

Boston, September 5th. A serious fire last night destroyed three houses in the center of the city. Seventeen families are without a home. The loss exceeds fifty thousand dollars. In rescuing a child, one of the firemen was badly burned about the hands and arms.

Two sec. after the reading is finished, remove the text, and ask *S:* "Tell me what you have been reading about." Record his report verbatim. Record also the type of his reading—whether (1) letter-by-letter, (2) by syllables, (3) hesitating, (4) straightforward, or (5) expressive.

S's report is graded in terms of the number of constituent ideas reproduced, in comparison with the full quota of 20 ideas, which is estimated as follows:

Three | houses | burned | Boston | September 5th | a serious fire | last night | destroyed | three buildings | in the center of the city | seventeen | families | are without a home | the loss exceeds | fifty thousand dollars | in rescuing | a child | one of the firemen | was badly | burned | about the hands and arms.

Count only correct reports, *e.g.*, the report—"A house on fire. A little boy had his hands burned"—counts only 3 credits.

Results.—(1) For the French text (53 words), the following average rates of reading were found: at 8 years, 45 sec., at 9 years, 40 sec., at 10 years, 30 sec., at 11 years, 25 sec.

(2) There is a positive relation between rate of reading and range of recall, *i.e.*, *S*'s who read slowly, recall fewer items. Even the slowest reader can report 2 ideas, but no *S* can report 6 ideas or more unless he is able to read the text within 1 min.

(3) Ability to read the text and recall at least 2 ideas is rare at the age of 7, but almost invariably present at the age of 8.

No. 30. Counting money.

Place upon a table, close together and in full view, two dimes and two nickels.[1] Ask *S*: "Count that money and tell me how much

[1] Goddard recommends the use of three 1-cent, and three 2-cent stamps. Binet and Simon used 6 *sous*, three '*simples*' and three '*doubles*.'

there is." "How much does that make"? The counting should not consume more than 10 sec., and the test is not passed unless it is completed within 15 sec. Note the manner of counting. Any error is termed a failure. The test can be passed by many children at 7 years, but by all children at 8 years.

No. 31. Naming four colors.

This is not a test of discrimination, but of naming. The child's ability to discriminate colors is nearly equal to the adult's.

METHOD.—Paste upon a piece of cardboard four rectangles (2 × 6 cm.) of red, yellow, blue, and green paper. Touch each piece, saying: "What color is that"? The test should be completed within 6 sec. A single error is interpreted as a failure.[1]

No. 32. Counting backwards from twenty.

Like the preceding, this is a test that depends in part upon the extent and nature of school or home training. Ask *S* to count from 20 to 0, backwards: if necessary, start him by saying, "20, 19, 18, what comes next"? To pass this test, *S* must complete the counting within 20 sec., and with not more than one error of omission or transposition.

Those who fail either balk entirely, or start correctly and then reverse and count forwards, or succeed only by the laborious method of finding each number by counting up from 1, and thus exceed the time-limit.

No. 33. Writing from dictation.

Require *S* to write with pen and ink, from dictation, the phrase: "The pretty little girls." As in No. 23, the writing must be sufficiently legible to be read by a person who did not know what was to be written. Only one-third can succeed at 7 years: all succeed at 8 years.

[1] The author has elsewhere (5) discussed the capacity of children to learn and employ color-names. Very much depends upon home and early school training.

No. 34. Differences between familiar objects recalled in memory. The test is identical with No. 16 of the 1st series. It is regarded as of special value, because it depends less upon school training than do the immediately preceding tests. The test is passed only if 2 of the 3 distinctions are correctly given within 2 min. Only one *S* in three succeeds at 6 years, nearly all succeed at 7 years, and all at 8 years.

G. TESTS FOR 9-YEAR OLD CHILDREN

No. 35. Knowing the date.

See if *S* knows: (1) the day of the week, (2) the month, (3) the day of the month, and (4) the year.

Children oftenest fail to know the year. The test is passed if the day of the month is given within 3 days, either way, of the real date. Capacity to pass the test is not commonly attained before the 9th year.[1]

No. 36. Reciting the days of the week.

Ask *S* to name the days of the week in order. This should be done in less than 10 sec., without error or hesitation.

No. 37. Making change.

This test is designed to afford insight into *S*'s capacity to perform a characteristic, every-day 'social' activity. It should be conducted in the form and spirit of a game, something in this way:

Ask *S* to 'play store' with you. Let him be 'store-man,' sell you some goods, and make change for you. Give him some boxes or other simple objects to sell, and supply him with an open box (about the size of a cigar-box), which contains 13 pennies, 5 nickels, 2 dimes, 2 quarters, 1 half-dollar, 1 silver dollar and 1 dollar bill. Offer to buy a box of him for four cents, saying: "I will pay you, say, four cents for one of your boxes, sha'n't I"? Then hand him

[1] Some of the younger school children tested by Binet and Simon had been given daily instruction in reciting the day, date in the month, etc., yet failed utterly to pass the test. This is cited as an instance of the uselessness of formal instruction prematurely given.

a quarter in exchange for the box, and extend your hand for the change.

The only 'pass' for this test is the correct counting out of the required 21 cents in change.[1] Record the manner, speed, and correctness of *S*'s counting. A few 7-year children, about one-third of 8-year children, and all 9-year children, pass the test.

No. 38. Definitions of familiar objects.

This is the same as No. 17: the child who is 9 years old is expected, however, to give definitions that shall contain something more than a simple statement of the functional nature of the term, and classificatory definitions should be given by many children, *e.g.*, "A horse is an animal to draw carriages." "A mamma is a woman that takes care of her children," etc.

No. 39. Reading and report.

In test No. 29, above described, the 9-year child should be able to report at least 6 items.

No. 40. Arrangement of five weights.

This is identical with No. 22 of the 1st series. The 9-year child should arrange all 5 weights correctly in 2 of 3 trials, and in not over 3 min. per trial.

H. TESTS FOR 10-YEAR OLD CHILDREN

No. 41. Reciting the months of the year.

Ascertain whether *S* can name in order the months of the year. To pass, the series must be given within 15 sec., and with not more than 1 error of omission or transposition.

[1] Goddard gives the child 25 pennies, 5 nickels, and 2 dimes, and hands him a quarter for a 9-cent article. The child must actually give 16 cents in change, as well as say it.

No. 42. Naming nine pieces of money.

Place on the table in a row, but not in regular order, a ten-dollar, a five-dollar, and a two-dollar bill, and six coins—dollar, half-dollar, quarter, dime, nickel, and cent. Let *S* point them out, one by one, and name them as he points.

No. 43. Sentence-building.

Follow the directions given in No. 26 of the 1905 series, but write the three terms, *girl, river, ball,* and supply *S* with pen and ink for the writing of his sentence, which should be finished within 1 min.

The responses to this test may be divided into (1) those in which there are virtually three distinct sentences, *e.g.*, "I saw a girl. Here is a river. Where is my ball?"; (2) those in which there appears a single sentence with two distinct ideas, *e.g.*, "The girl lives near a river and has a ball"; (3) those in which the three terms are combined into a single idea, *e.g.*, "The girl lost her ball in the river"; and (4) those in which the three terms are combined in a more elaborately contrived, well coordinated sentence, *e.g.*, "One day, when I was a little girl, I was walking along the bank of a river, when I chanced to look up and spied a ball floating on the water." Replies of the first type are classed as failures: on this basis it is found that the test can be passed by a few 8-year old *S*'s, by about one-third at 9, one-half at 10. The good sense, as well as the grammatical construction of the sentence, should be considered in estimating *S*'s intellectual capacity.

No. 44. Replies to problem-questions.

A test that can be passed by about one-half of 10-year old *S*'s is formed by selecting two series of questions from No. 27 of the 1905 series, as follows: Series 1: Questions 6, 10, 7; Series 2, Questions 3, 15, 22, 18, 21. *S* should be allowed at least 20 sec., if desired, to formulate his answer for each question. His rank is estimated on the basis of the complete series of answers, and *E* may 'tolerate' 2 'bad' answers in 5. As not more than half of 10-year old *S*'s can answer the majority of Series 2 satisfactorily, the test forms a transition between the 10-year and the 11-year tests, proper.

I. TESTS FOR 11-YEAR OLD CHILDREN

No. 45. Detecting absurd or contradictory statements.

Alienists sometimes, in the course of the examination of patients, endeavor to see whether the patients will assent to ridiculous assertions. The present test is designed for a similar purpose, but the inquiry is so framed as to challenge the pupil to detect and explain the absurdity.

Explain to *S:* "I'm going to tell you something that has a 'catch' in it: you listen carefully and see if you can tell me where the nonsense is."

(1) "An unfortunate bicycle rider broke his head and died instantly: he was picked up and carried to a hospital, and they do not think that he will recover."

(2) "I have three brothers, Paul, Ernest, and myself."

(3)[1] "John is taller than I am; Henry is taller than John, and I am taller than Henry."

(4) "We met a man who was finely dressed: he was walking along the street with his hands in his pockets, and twirling his cane."

(5) "The engineer said that the more cars he had on his train, the faster he could go."

(6) "The other day I was walking to Boston when I met a fine team: there were two men on the front seat, and a man, a woman, and a baby on the back seat: so five were going to Boston."

The test is passed if the absurdity of 3 of the first 5 statements is detected. Almost no child can do this at 9 years, scarcely one child in four at 10 years, and about one child in two at 11 years.

No. 46. Sentence-building.

Test No. 43, when tried by 11-year old *S*'s, should yield sentences of the 3d or 4th type.

[1] Nos. 3–6 are substituted for the rather blood-curdling series proposed by Binet and Simon, viz: (3) "There was found in the park today the body of an unfortunate young girl, frightfully mutilated, and chopped into 18 pieces. It is thought that she committed suicide." (4) "There was an accident on the railroad today, but it was not serious; there were only 48 killed." (5) "Some one said: 'If I kill myself in despair some day, I sha'n't choose Friday to do it, because Friday brings you bad luck.'"

No. 47. Naming words.[1]

Require *S* to name as many words as he can in 3 min. Explain that any words will do, and stir him to do his best by assuring him that some children can name 200 words in this time. If necessary, start a series for him, *e.g.*, "beard, table, skirt, carriage," etc. If possible, record the series given by *S*. Count the number at least.

To pass the test, *S* must name at least 60 words. Bright children can name from 100 to 200: one child is recorded who named 218. The performance of the child often throws light upon his intellectual equipment and general alertness. Clever children hit upon series of related terms, especially of abstract and descriptive terms.

No. 48. Definition of abstract terms.[2]

Ask *S:* "What is charity"? "What is justice"? "What is kindness"? The test is passed if two of the three terms are defined acceptably. Most 11-year old children can pass the test, about one-third of 10-year old children, but only a few can construct definitions at 8 or 9 years of age.

The definition of *charity* should contain two ideas—that of unfortunates and that of the good that is done them: the definition of *justice* should contain either the notion of law or that of treating persons according to their merits: the definition of *kindness* should contain the notion of affection, sentiment, tenderness, assistance, etc.

No. 49. Arranging words in a sentence.[3]

This test differs from No. 43 in that all the parts of the sentence are supplied, and *S* has only to piece these together to 'make sense.' A time-limit of 1 min. is placed upon each sentence: very bright *S*'s may effect the arrangement in 5 sec.

Copy upon three small sheets of paper, with a typewriter or in legible script, the three mutilated sentences, following precisely

[1] See Chapter IX, Test 33, for results of similar tests when applied to adults.
[2] Cf. No. 30 of the 1st, and No. 56 of the present series.
[3] Cf. Test 47, Chapter XI.

the arrangement here given. Hand S one sheet at a time and ask him to make a correct sentence (orally) from the words on the sheet.

(1) a defends
dog good his
master courageously

(2) my have teacher
I the correct
asked paper to

(3) home our early
we country in left
visit the to friends

The corrected sentences are: (1) A good dog defends his master courageously. (2) I have asked the teacher to correct my paper. (3) We left home early to visit our friends in the country.

J. TESTS FOR 12-YEAR OLD CHILDREN

No. 50. Memory for seven digits.

This is the same as No. 3. Let *S* know that there are to be 7 digits. If he repeats all 7 correctly in 1 of 3 trials (3 different numbers), the test is passed.

No. 51. Rimes.

Use No. 24 in the 1905 series to see if *S* can make at least 3 rimes in 1 min.

No. 52. Memory for sentences (26 syllables).

Children 12 years old should be able to repeat the first 13 sentences of No. 2.

No. 53. Problems.

This is another so-called 'common-sense' test.

(1) Ask *S* to supply the missing word in the following: "A person who was walking in the woods in the park stopped in fright and

ran to the nearest policeman, saying that she had just seen hanging from the branch of a tree a ——."

(2) "My neighbor has been having queer visitors: first a doctor, then a lawyer, then a priest. What's happening there"?

To pass the test, both 'problems' must be solved.

K. TESTS FOR 13-YEAR OLD CHILDREN.

No. 54. Drawing from a design cut in a quarto-folded paper.

Follow the directions in No. 29 of the 1905 series. If *S* succeeds, ask him if he has tried the test before.

No. 55. Juxtaposed triangles.

Select a card of about twice the size of that represented in Fig. 64. Cut it in two parts diagonally from *a* to *b*. Place the card with the two halves contiguous, in *S*'s view on a table, and say to him: "Look carefully at this card, especially at this [lower] half. Suppose we should turn this half upside down, and place this corner [*c*] touching this point [*b*] so that this edge [*bc*] shall touch this edge [*ba*], what would the whole figure look like then? Now, I'm going to pick up this lower half. I want you to imagine it turned over and laid up against the upper half as I have said. Draw the

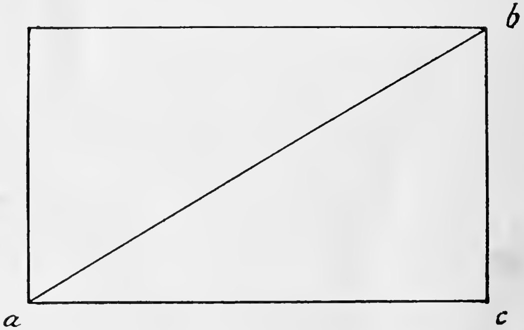


FIG 64. THE TEST OF JUXTAPOSED TRIANGLES.
(No. 55, 1908 Binet-Simon series.)

whole figure for me, as it would look then. Begin with the upper half that you see before you."

The test is difficult: the essential point for success is to preserve the right angle *bca*, and to make *cb* shorter than *ba*.

No. 56. Distinctions between abstract terms.

This test is similar to No. 30 of the 1905 series, except in the choice of terms. Ask *S*: (1) "What is the difference between *pleasure* and *happiness*"? (2) "What is the difference between *evolution* and *revolution*"? (3) "What is the difference between *poverty* and *misery*"? (4) "What is the difference between *pride* and *pretention*"?

Treatment of Data.—In handling these numerous tests in serial arrangement for diagnostic purposes, the adoption of a simple system of recording the data for the tests will be found helpful. The following suggestions are made to this end. Prepare a large record sheet on which the numbers and descriptive titles of the tests are spread out horizontally (somewhat as in Table 99). Divide the array of tests by vertical lines in accordance with the division by age-norms as already indicated. Enter the results for each child under the several tests, using convenient symbols, *e.g.*, *Y* for complete success, *S* for partial success, *P* for poor outcome, *O* for complete failure. It will then be possible and profitable to add a second symbol to indicate the nature or cause of the failure, *e.g.*, *A* for a failure characterized by some absurdity that indicates lack of judgment or common-sense, *I* for failure through ignorance, *i.e.*, inability to comprehend what is wanted or to supply the answer, *R* for failure through resistance, stubbornness, etc., *T* for failure through timidity, *D* for failure through defective attention (accidental distraction, preoccupation, scatter-brained condition, etc). Partial success may also often be indicated by the use of a subjoined fraction (*e.g.*, $\frac{1}{2}$) to indicate the proportion of the test performed correctly.

When *S*'s results have been distributed in this manner, *E* is next confronted with the problem of determining whether these results indicate normal, retarded, or advanced mental development. Since there can, in the nature of the case, be no hard and

fast lines, as there is no single, decisive test of intelligence, and since most *S*'s who fail at a given test succeed with some of the tests of the next grade of difficulty (see Table 99), it follows that the determination of mental status must be based upon some quasi-artificial rule.. Binet and Simon have found the following rules most serviceable in practise: (1) The intellectual status of a given *S* is that of the oldest age at which he can pass successfully all of the tests save one that are assigned to the age in question. But (2) to the intellectual level thus determined, *E* may add one year for every 5 tests that are passed above the year in question. A child whose final rank places him 3 years behind the norm of his age is to be considered mentally defective.[1]

In illustration: a boy 9 years old passed only 3 tests in that year: on the basis of Rule 1, his status is that of an 8-year child, or one year of retardation. But the same boy also passed 3 of the 10-year tests: hence he passed in all 6 tests above the 8-year limit, and therefore by Rule 2, is entitled to rank "at age," or normal. In exceptional cases, this rule may necessarily be abrogated, or it may be modified at *E*'s discretion.

Results.—(1) In the examination of 203 French children, aged 3–12 years, Binet and Simon found that 103 children ranked as of normal capacity: 42 were advanced 1 year; 2 were advanced 2 years; 44 were retarded 1 year; 12 were retarded 2 years: in other words, about one-half were "at age," 44 were advanced, and 56 were retarded.[2]

(2) Children selected by their teachers as possessing unusual ability invariably rank as 'advanced' in these tests: again, an examination of 14 children who were 3 years behind in school grades showed that all were retarded in these tests by from 1 to 6 years (averaging 2.5 years), Both of these facts confirm the value of the series of tests.

(3) Decroly and Degand have applied the 1908 series in Brussels to 45 normal children from well-to-do families. These writers

[1]For illustrations of the application of the plan of rating the intelligence of a feeble-minded child in terms of the intelligence of a normal child at different ages, see Binet and Simon (2).

[2]The fact that only 14 of 203 children varied by as much as 2 years from the normal rating, demonstrates, in the opinion of the authors, that intellectual capacity, as measured by these tests, shows less variation than do such anthropometric values as height, weight, skull-capacity, etc.

find the series valuable as a whole, but they regret the omission of certain of the 1905 tests which seemed to them of special value, and they do not find that Binet and Simon have been entirely successful, either in eliminating the effect of accidental school and home training, or in grading the tests correctly in respect to the several age-norms. Their critique is summarized in Table 100. It will be seen that there are, in substance, six points of criticism. (*a*) Certain tests are too simple for the age to which they have been assigned. (*b*) Certain tests (in the 13th year) are too difficult for the assigned year. (*c*) Certain tests are too 'mechanical,' in that they may be satisfied by a quasi-automatic form of response that does not compel the display of real intellectual ability.[1] (*d*) Certain tests are too 'schooly' (*trop scolaire*), *i.e.*, their outcome is to a greater or less degree contingent upon *training* that the child may have received either at home or at school. (*e*) The tests of rote memory should be graded to supply norms of performance, now lacking, between the ages 7 and 12. (*f*) The sentence-building tests (Nos. 43 and 46) would be more valuable if they could be arranged to permit verbal execution within a prescribed time-limit.

TABLE 100

Critique of the 1908 Binet-Simon Series. (Decroly and Degand)

NO.	NAME	CRITICISM
1	Comprehension of language	Can be done earlier than 3d year
5	Family name	Training
6	Own sex	Training
7	Naming familiar objects	Can be done earlier than 4th year
13	Counting four pennies	Training
16	Elementary esthetic judgment	Can be done at the 3d year
17	Defining familiar objects	Can be done at the 3d year
18	Executing a triple order	Can be done at the 3d year
19	Knowing own age	Training
22	Number of fingers	Training
23	Writing from copy	Training
25	Memory for digits	Norms needed between 7th and 12th year

[1] Of these, Nos. 36 and 41—reciting the days of the week and the months of the year—could, as Decroly and Degand suggest, be made non-mechanical by asking the child to recite the days or the months backward, or by asking him to name the day or the month preceding a given day or month.

NO.	NAME	CRITICISM
27	Counting three pennies	Too mechanical
28	Naming four coins	Training
29	Reading and report	Training
30	Counting money	Training
31	Naming four colors	Can be done at the 4th year
32	Counting backwards	Training
33	Writing from dictation	Training
34	Differentiating from memory	Can be done earlier than the 8th year
36	Reciting days of the week	Mechanical. Could be modified
39	Reading and report	Training, at least for some children
40	Arranging five weights	Can be done at 5th or 6th year
41	Reciting the months	Mechanical. Could be modified
43	Sentence building	Verbal execution should be permitted
45	Detecting absurdities	Can be done at the 6th year
46	Sentence building	As 43
47	Naming words	Can be done earlier than the 11th year
50	Memory for digits	See No. 25
54	Drawing a cut paper	Too difficult for 13th year
55	Juxtaposed triangles	Too difficult for 13th year
56	Distinction abstract terms	Too difficult for 13th year

NOTES.—While these tests have been termed "a measuring scale of intelligence," it is not to be supposed that they have all the neatness and precision of the chemist's balance, or even of the calipers and tape of the anthropometrist. The criticisms of Decroly and Degand are sufficient evidence that we must do preliminary work in standardizing our tests before we can hope to rate the individual child with any accuracy.

Furthermore, totally inexperienced E's can not expect to apply these tests with success. The elimination of delicate apparatus does not eliminate the need of skill and nicety in the administration of tests, for the child's nervous system is more delicate than any instrument of precision.

Particularly in applying these tests to the subnormal child, is the attitude of the examiner all-important. To quote from Goddard: "Always the child must be won. Sometimes it is easy, sometimes it is difficult. The questioner should be very tactful and careful until he sees that the child is at ease . . . At

all events, *get down to the level of the child.* Never tell a child his answer is wrong. Always encourage. . . . On the other hand, do not insist that he respond, just because it seems to you that he must know. . . . While examining the child, forget all your preconceived ideas. Regard him as an unknown quantity, an x that is to be determined."

REFERENCES

(1) A. Binet and T. Simon, Le développement de l'intelligence chez les enfants, in A. P., 14: 1908, 1-94.

(2) A. Binet and T. Simon, L'intelligence des imbéciles, in A. P., 15: 1909, 1–147.

(3) O. Decroly and (Mlle.) J. Degand, La mesure de l'intelligence chez des enfants normaux d'après les tests de Mm. Binet et Simon: nouvelle contribution critique, in Ar. P., 9: 1910, 81–108.

(4) H. H. Goddard, A measuring scale for intelligence, in The Training School, 6: January, 1910, 146–155.

(5) Dr. and Mrs. G. M. Whipple, The vocabulary of a three-year-old boy, with some interpretative comments, in Pd. S., 16: 1909, 1–22, especially p. 19.

LIST OF MATERIALS

Roman numerals refer to test-numbers, italicized numerals to page-numbers. Items starred refer to materials that are recommended, but not prescribed, or to materials for the conduct of alternative or supplementary tests.

The Materials may be ordered of C. H. Stoelting Company, 121 N. Green St., Chicago, Illinois, who will quote prices on application.

I. SPECIAL APPLIANCES

Adding machine, *31**.
Acoumeter, Lehmann's, 18.
Analyzer, Sommer's tridimensional, 13*.
Astigmatic chart, Verhoeff's, 14.

Balls, five wooden, 52.

Card of objects, Binet's, 32.
Caliper, Vernier, 1.
Calipers, head, 3.
Color-blindness apparatus, Hering's 16*.
Color-blindness cards, Nagel's, 16.
Color-blindness worsteds, Holmgren's, 16.
Colored papers, set of four, 54.
Conformateur, 3*.
Counter, mechanical, 36*.
Cubes, wooden, 52.

Discrimination of brightness, Whipple's apparatus for, 17.
Discrimination of grays, Whipple's apparatus for, 17.
Dynamometer, back and leg, 7, 8.
Dynamometer, Smedley's, 6, 9.

Ergograph, Mosso's, 9*.
Esthesiometer, Jastrow's, 23.

Fork, Blake's, 18.
Fork, fifty-vibration, 24*.
Forks for pitch discrimination, 19.

Hammer, soft tipped, 19.

Key, telegraph, 10*, 13.
Krypteon, 25*.
Kymograph, and accessories, 9, 10, 13, 24*, 30*, 42.

Lenses, trial, 14.

Maddox rod, 15.
Memory apparatus, Jastrow's, 38*.
Mouth-pieces for spirometer, 5.

Parallelopipeds, two wooden, 52.
Pencil, electrical, 35*.
Pendulum, second's, 10, 21, 22, 25*, 38*.
Picture, Australians, 32; Hindoos, 31; The Disputed Case, 32; The Orphans' Prayer, 32*; Washington and Sally, 32*.
Pictures, colored, set of eight, 53, 54.
Pictures, colored, group of twelve on cardboard, 53.
Pressure-pain balance, Whipple's, 21, 22.
Prism, twenty-degree, 36*.
Prisms, set of four, 15.
Pyramids, three wooden, 52.

Scales, anthropometric, 2.
Signal-magnet, see time-marker.
Smoking stand, see kymograph.
Spirometer, 5.
Stadiometer, 1.
Steadiness tester, Whipple's, 13.
Stenopaic lens, Stevens', 15*.
Suggestion blocks, 40.

II. SPECIAL PRINTED FORMS

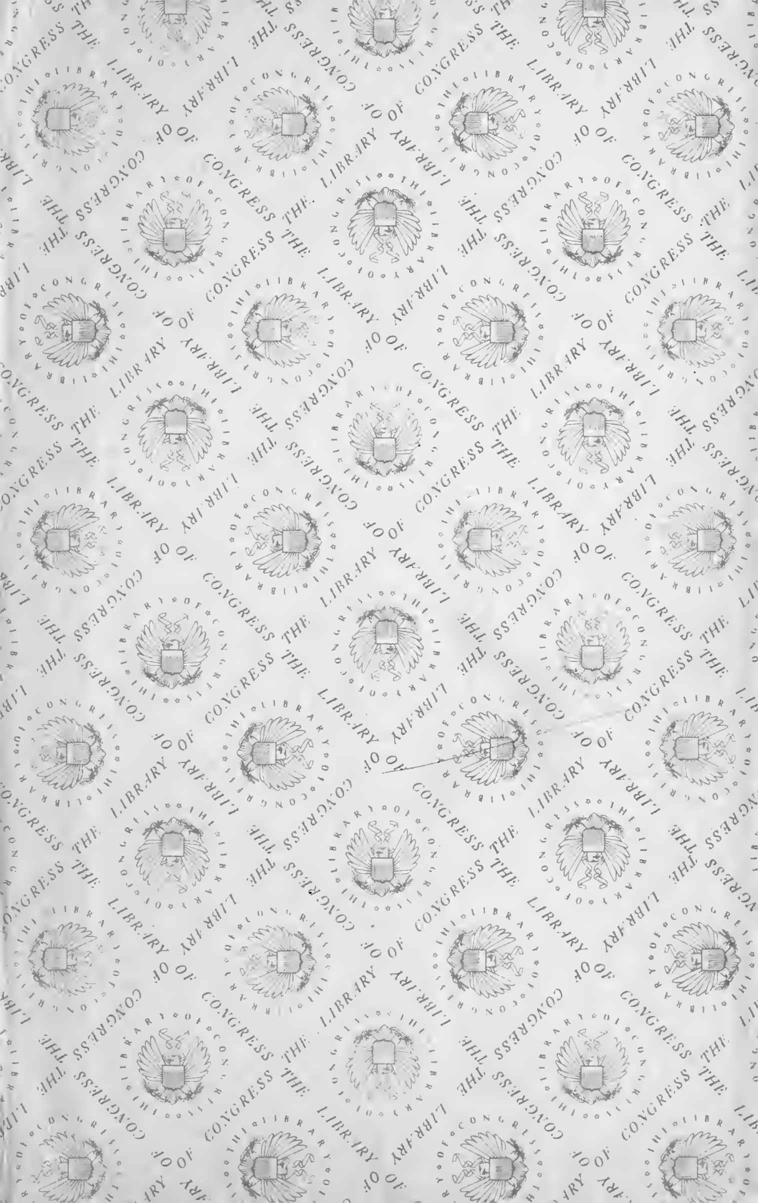
III. GENERAL APPLIANCES AND MATERIALS

INDEX OF NAMES

Roman numerals refer to test-numbers, italicized numerals to page-numbers.

INDEX OF SUBJECTS

For authors quoted, see Index of Names; for apparatus, see List of Materials.


THE LIBRARY OF CONGRESS